

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:23:09 ; Search time 222.882 Seconds
(without alignments)

514.627 Million cell updates/sec

Title: US-10-624-714-1

Perfect score: 27
Sequence: 1 gtcgaacagactttcccaaacagac 27

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002s:*
- 7: Geneseqn2003as:*
- 8: Geneseqn2003bs:*
- 9: Geneseqn2003cs:*
- 10: Geneseqn2004s:*

Fred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	100.0	2037	4	Aah51958 Mycobacte
2	27	100.0	2037	7	Aca40280 Prokaryot
3	27	100.0	2466	7	Aca38623 Prokaryot
4	27	100.0	110000	4	Aai99682 Mycobacte
5	27	100.0	110000	4	Aai99683 Mycobacte
6	20.6	76.3	597	7	Abz38859 N. gonorr
7	20.6	76.3	699	7	Abz38855 N. gonorr
8	20.6	76.3	35133	3	Aaa81518 N. mening
9	20.6	76.3	110000	3	Continuation (4 of
10	20.6	76.3	34980	3	Aaf21611 Neisseria
11	20.2	74.8	6862	4	Aas46239 Tumour su
12	20.2	74.8	6862	6	AbL32222 Human imm
13	20.2	74.8	6862	6	Aas61081 Human gen
14	19	70.4	1728	7	Aca45661 Prokaryot
15	18.6	68.9	13169	4	AbL18727 Drosophil
16	18.6	68.9	35832	4	AbL18726 Drosophil
17	18	66.7	505	6	Abq43857 Oligonucl
18	18	66.7	505	6	Abq43856 Oligonucl
19	18	66.7	584	6	Abq42326 Oligonucl
20	18	66.7	584	6	Abq42327 Oligonucl
21	18	66.7	680	3	Aaf13731 Aspergill
22	18	66.7	15224	6	Abk31359 signal tr
23	18	66.7	15224	6	AbL70318 Chemical

ALIGNMENTS

RESULT 1

AAH51958
ID AAH51958 standard; DNA; 2037 BP.

XX AC AAH51958;

XX DT 04-SEP-2001 (first entry)

XX DE Mycobacterium tuberculosis potential drug target gene SEQ ID 12.

XX KW Drug target; growth; organism viability; characterisation; ds.

XX OS Mycobacterium tuberculosis.

XX FN WO200135317-Al.

XX PD 17-MAY-2001.

XX PF 13-NOV-2000; 2000WO-US0311152.

XX PR 12-NOV-1999; 99US-0165086P.

XX PR 12-NOV-1999; 99US-0165124P.

XX PR 01-FEB-2000; 2000US-0179531P.

XX (REGC) UNIV CALIFORNIA.

XX Eisenberg D, Rotstein SH, Marcotte EM;

XX WPI; 2001-329193/34.

XX P-PSDB; AAG81107.

XX Identifying nucleotide or polypeptide sequence for use as drug target,
XX involves providing algorithm that analyzes a functional relationship
XX between nucleotide or polypeptide sequences, and comparing the sequences.

XX Disclosure; Page 55; 207pp; English.

XX This invention relates to a method for identifying a nucleotide or
XX polypeptide sequence that may be a drug target, or essential for growth
XX or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
XX represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium
XX tuberculosis proteins which are potential drug targets. The DNA and
XX protein sequences are used to illustrate the method of the invention. The
XX method involves providing an unknown nucleotide or polypeptide sequences,
XX and comparing it to a number of sequences along with at least one
XX algorithm capable of analysing a functional relationship between

Ras61263 Human gen
Aas91050 DNA encod
AbL09937 Drosophil
AbL09936 Drosophil
Abq48693 Oligonucl
Abq48692 Oligonucl
Abq41462 Oligonucl
Abq41463 Oligonucl
Aah19473 Lactococc
AbL13398 Drosophil
Continuation (7 of
Continuation (8 of
Abv98614 Human pan
Abq36792 Oligonucl
Abq36793 Oligonucl
Aca30791 Prokaryot
Aak70288 Human imm
Aak70289 Human imm
Continuation (3 of
Adb55631 Toxicity-
Aaa69941 Human ova
Abn72835 Ovarian c

CC nucleotide and polypeptide sequences. The method is useful for
 CC characterizing the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism
 XX
 SQ Sequence 2037 BP; 396 A; 639 C; 651 G; 351 T; 0 U; 0 Other;
 Query Match 100.0%; Score 27; DB 4; Length 2037;
 Best Local Similarity 100.0%; Pred. No. 0.0079;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 GTCGACGAGACTTCCCAACCGAC 27
 Db 1821 GTCGACGAGACTTCCCAACCGAC 1847
 RESULT 2
 ACA40280
 ID ACA40280 standard; DNA; 2037 BP.
 XX
 AC ACA40280;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Prokaryotic essential gene #21937.
 XX
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR P-PSDB; ABU36410.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 14; SEQ ID NO 28150; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *X. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2037 BP; 396 A; 639 C; 651 G; 351 T; 0 U; 0 Other;
 Query Match 100.0%; Score 27; DB 7; Length 2037;
 Best Local Similarity 100.0%; Pred. No. 0.0079;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCGACGAGACTTCCCAACCGAC 27
 Db 1821 GTCGACGAGACTTCCCAACCGAC 1847
 RESULT 3
 ACA38623
 ID ACA38623 standard; DNA; 2466 BP.
 XX
 AC ACA38623;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Prokaryotic essential gene #20280.
 XX
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 XX
 OS Mycobacterium bovis.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR P-PSDB; ABU34753.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 14; SEQ ID NO 26493; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway;
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcr_sequences

XX
 SQ Sequence 2466 BP; 454 A; 808 C; 800 G; 404 T; 0 U; 0 Other;

Query Match 100.0%; Score 27; DB 7; Length 2466;
 Best Local Similarity 100.0%; Pred. No. 0.0081;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCGAACGAGACTTCCCAACCGAC 27
 DB 2247 GTCGAACGAGACTTCCCAACCGAC 2273

RESULT 4

WP sequence split into 45 fragments LOCUS: AAI99682 Accession Aai99682

Fragment Name	Begin	End
WP AAI99682_00	1	110000
WP AAI99682_01	100001	210000
WP AAI99682_02	200001	310000
WP AAI99682_03	300001	410000
WP AAI99682_04	400001	510000
WP AAI99682_05	500001	610000
WP AAI99682_06	600001	710000
WP AAI99682_07	700001	810000
WP AAI99682_08	800001	910000
WP AAI99682_09	900001	1010000
WP AAI99682_10	1000001	1110000
WP AAI99682_11	1100001	1210000
WP AAI99682_12	1200001	1310000
WP AAI99682_13	1300001	1410000
WP AAI99682_14	1400001	1510000
WP AAI99682_15	1500001	1610000
WP AAI99682_16	1600001	1710000
WP AAI99682_17	1700001	1810000
WP AAI99682_18	1800001	1910000
WP AAI99682_19	1900001	2010000
WP AAI99682_20	2000001	2110000
WP AAI99682_21	2100001	2210000
WP AAI99682_22	2200001	2310000
WP AAI99682_23	2300001	2410000
WP AAI99682_24	2400001	2510000
WP AAI99682_25	2500001	2610000
WP AAI99682_26	2600001	2710000
WP AAI99682_27	2700001	2810000
WP AAI99682_28	2800001	2910000
WP AAI99682_29	2900001	3010000
WP AAI99682_30	3000001	3110000

WP AAI99682_31 3100001 3210000
 WP AAI99682_32 3200001 3310000
 WP AAI99682_33 3300001 3410000
 WP AAI99682_34 3400001 3510000
 WP AAI99682_35 3500001 3610000
 WP AAI99682_36 3600001 3710000
 WP AAI99682_37 3700001 3810000
 WP AAI99682_38 3800001 3910000
 WP AAI99682_39 3900001 4010000
 WP AAI99682_40 4000001 4110000
 WP AAI99682_41 4100001 4210000
 WP AAI99682_42 4200001 4310000
 WP AAI99682_43 4300001 4410000
 WP AAI99682_44 4400001 4411529
 ID AAI99682 standard; DNA; 4411529 BP.
 XX
 AC AAI99682;
 XX
 DT 15-JAN-2002 (first entry)
 XX
 DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
 XX
 KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
 XX variation; epidemiology; patient treatment; epidemic monitoring; ds.
 OS Mycobacterium tuberculosis.
 XX
 FN US6294328-B1.
 XX
 PD 25-SEP-2001.
 XX
 PF 24-JUN-1998; 98US-00103840.
 XX
 PR 24-JUN-1998; 98US-00103840.
 XX
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Fleischmann RD, White OR, Fraser CM, Venter JC;
 XX WPI; 2001-647261/74.
 DR
 XX
 PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
 PT determining the nucleotide sequence of the strain at positions in the
 PT genome corresponding to positions where *M. tuberculosis* strains CDC 1551
 XX and H37Rv differ.
 PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
 XX
 CC The invention relates to evaluating strain variation within and between
 CC different populations of the tuberculosis bacterial pathogen,
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the
 CC nucleotide sequence of the first strain at positions in the complete
 CC sequence of the genome that correspond to positions that differ in the
 CC nucleotide sequences of *M. tuberculosis* strains CDC 1551 (AAI99683) and
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
 CC *M. tuberculosis* and has valuable application in the fields of
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic
 CC monitoring. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from USPTO at seqdata.uspto.gov/sequence.html?DocID=629432891
 XX
 SQ Sequence 4411529 BP; 758565A; 1449983C; 1444602G; 758379T; 0U; 0Other;

Query Match 100.0%; Score 27; DB 4; Length 110000;

Best Local Similarity 100.0%; Pred. No. 0.014;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTCGAACGAGACTTCCCAACCGAC 27

DB 55481 GTCGAACGAGACTTCCCAACCGAC 55507

RESULT 5

AAI99683_00
 WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683
 WP Fragment Name Begin End
 WP AAI99683_00 1 110000
 WP AAI99683_01 100001 210000
 WP AAI99683_02 200001 310000
 WP AAI99683_03 300001 410000
 WP AAI99683_04 400001 510000
 WP AAI99683_05 500001 610000
 WP AAI99683_06 600001 710000
 WP AAI99683_07 700001 810000
 WP AAI99683_08 800001 910000
 WP AAI99683_09 900001 1010000
 WP AAI99683_10 1000001 1110000
 WP AAI99683_11 1100001 1210000
 WP AAI99683_12 1200001 1310000
 WP AAI99683_13 1300001 1410000
 WP AAI99683_14 1400001 1510000
 WP AAI99683_15 1500001 1610000
 WP AAI99683_16 1600001 1710000
 WP AAI99683_17 1700001 1810000
 WP AAI99683_18 1800001 1910000
 WP AAI99683_19 1900001 2010000
 WP AAI99683_20 2000001 2110000
 WP AAI99683_21 2100001 2210000
 WP AAI99683_22 2200001 2310000
 WP AAI99683_23 2300001 2410000
 WP AAI99683_24 2400001 2510000
 WP AAI99683_25 2500001 2610000
 WP AAI99683_26 2600001 2710000
 WP AAI99683_27 2700001 2810000
 WP AAI99683_28 2800001 2910000
 WP AAI99683_29 2900001 3010000
 WP AAI99683_30 3000001 3110000
 WP AAI99683_31 3100001 3210000
 WP AAI99683_32 3200001 3310000
 WP AAI99683_33 3300001 3410000
 WP AAI99683_34 3400001 3510000
 WP AAI99683_35 3500001 3610000
 WP AAI99683_36 3600001 3710000
 WP AAI99683_37 3700001 3810000
 WP AAI99683_38 3800001 3910000
 WP AAI99683_39 3900001 4010000
 WP AAI99683_40 4000001 4110000
 WP AAI99683_41 4100001 4210000
 WP AAI99683_42 4200001 4310000
 WP AAI99683_43 4300001 4403765
 ID AAI99683 Standard; DNA; 4403765 BP.
 AC AAI99683;
 AT 15-JAN-2002 (first entry)
 DT Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
 DE Mycobacterium tuberculosis; strain H37Rv; genome; ds.
 KW Mycobacterium tuberculosis; strain H37Rv; genome; ds.
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
 XX Mycobacterium tuberculosis.
 XX US6294328-B1.
 XX 25-SEP-2001.
 XX 24-JUN-1998; 98US-00103840.
 XX 24-JUN-1998; 98US-00103840.
 XX (GENO-) INST GENOMIC RES.
 XX Fleischmann RD, White OR, Fraser CM, Venter JC;
 XX WPI; 2001-647261/74.
 DR

XX Evaluating strain variation of Mycobacterium tuberculosis, comprises
 PT determining the nucleotide sequence of the strain at positions in the
 PT genome corresponding to positions where M. tuberculosis strains CDC 1551
 PT and H37Rv differ.
 XX Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
 PS The invention relates to evaluating strain variation within and between
 CC different populations of the tuberculosis bacterial pathogen,
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the
 CC nucleotide sequence of the first strain at positions in the complete
 CC sequence of the genome that correspond to positions that differ in the
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
 CC M. tuberculosis and has valuable application in the fields of
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic
 CC monitoring. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from USPTO at seqdata.uspto.gov/sequence.html?DocID=629432881
 XX SQ Sequence 4403765 BP; 757105A; 1447799C; 1441301G; 757371T; 0U; 1890Other;
 Query Match 100.0%; Score 27; DB 4; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 0.014;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GTCGACGAGACTTCCCAACCGAC 27
 Db 55423 GTCGACGAGACTTCCCAACCGAC 55449
 RESULT 6
 ABZ38869/c
 ID ABZ38869 standard; DNA; 597 BP.
 XX AC ABZ38869;
 XX DT 07-MAR-2003 (first entry)
 XX DE N. gonorrhoeae nucleotide sequence SEQ ID 2327.
 XX KW Antibacterial; infection; vaccine; gene therapy; gene; ds.
 XX OS Neisseria gonorrhoeae.
 XX PN WO200279243-A2.
 XX PD 10-OCT-2002.
 XX PF 12-FEB-2002; 2002WO-IB002069.
 XX PR 12-FEB-2001; 2001GB-00003424.
 XX PA (CHIR-) CHIRON SPA.
 XX PI Fontana MR, Pizza M, Massignani V, Monaci E;
 XX WPI; 2003-058415/05.
 XX DR P-PSDB; ABP77899.
 XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
 XX medicament for treating or preventing N. gonorrhoeae infection.
 XX PS Disclosure; Page 357; 815pp; English.
 CC The present invention relates to proteins from Neisseria gonorrhoeae.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing N. gonorrhoeae
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid

CC molecules of the invention
XX SQ Sequence 597 BP; 137 A; 134 C; 196 G; 130 T; 0 U; 0 Other;
Query Match 76.3%; Score 20.6; DB 7; Length 597;
Best Local Similarity 85.2%; Pred. No. 7.3;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 GTCGAACGAGACTTCCCAACCGAC 27
Db 114 GTCGAACGAGACTTCCCAACCGAC 88
RESULT 7
ABZ38855
ID ABZ38855 standard; DNA; 699 BP.
XX AC ABZ38855;
XX DT 07-MAR-2003 (first entry)
XX DE N. gonorrhoeae nucleotide sequence SEQ ID 2299.
XX KW Antibacterial; infection; vaccine; gene therapy; gene; ds.
XX OS Neisseria gonorrhoeae.
XX PN WO200279243-A2.
XX PD 10-OCT-2002.
XX PF 12-FEB-2002; 2002WO-IB002069.
XX PR 12-FEB-2001; 2001GB-00003424.
XX PA (CHIR-) CHIRON SPA.
XX PI Fontana MR, Pizza M, Massignani V, Monaci E;
XX DR WPI; 2003-058415/05.
XX DR P-PSDB; ABP77885.
XX PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a
XX PT medicament for treating or preventing N. gonorrhoeae infection.
XX PS Disclosure; Page 354; 815pp; English.
XX CC The present invention relates to proteins from Neisseria gonorrhoeae.
XX CC Also disclosed are the nucleic acid molecules encoding the proteins and
XX CC antibodies that specifically bind to the proteins. The composition
XX CC comprising the protein, nucleic acid or antibody is useful for the
XX CC manufacture of a medicament for treating or preventing N. gonorrhoeae
XX CC infection, this may be in the form of a vaccine or gene therapy.
XX CC Sequences given in records ABZ37706-ABZ42016 represent nucleic acid
XX CC molecules of the invention
XX SQ Sequence 699 BP; 155 A; 227 C; 162 G; 155 T; 0 U; 0 Other;
Query Match 76.3%; Score 20.6; DB 7; Length 699;
Best Local Similarity 85.2%; Pred. No. 7.5;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 GTCGAACGAGACTTCCCAACCGAC 27
Db 490 GTCGAACGAGACTTCCCAACCGAC 516
RESULT 8
AAA81518
ID AAA81518 standard; DNA; 35133 BP.
XX AC AAA81518;
XX

DT 04-DEC-2000 (first entry)
XX N. meningitidis partial DNA sequence gnm_65 SEQ ID NO:65.
XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX OS Neisseria meningitidis.
XX PN WO200022430-A2.
XX PD 20-APR-2000.
XX PF 08-OCT-1999; 99WO-US023573.
XX PR 09-OCT-1998; 98US-0103794P.
XX PR 30-APR-1999; 99US-0132068P.
XX (CHIR) CHIRON CORP.
XX PA Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
XX PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
XX PI Rappuoli R, Pizza M;
XX DR WPI; 2000-318079/27.
XX PT Isolated nucleotide sequences of Neisseria meningitidis which can be used
XX PT in the diagnosis and treatment of N. meningitidis infection and other
XX PT Neisserial infections, for example, N.gonorrhoea.
XX PS Claim 7; Page 1396-1406; 1760pp; English.
XX CC The present invention describes methods of obtaining immunogenic proteins
XX CC from Neisseria genomic sequences. AAA81453 to AAA82414 represent
XX CC specifically claimed Neisseria meningitidis genomic DNA sequences;
XX CC AAA81260 to AAA81303 and AAB25620 to AAB25663 represent Neisseria DNA
XX CC sequences and their corresponding proteins; AAA81254 to AAA81259 and
XX CC AAA81304 to AAA81321 represent PCR primers used in the isolation of
XX CC Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent
XX CC Neisseria meningitidis MenB polynucleotide ORF sequences, which are all
XX CC used in the exemplification of the present invention. The nucleic acid
XX CC sequences, protein sequences, and antibodies against them, can be used in
XX CC the manufacture of a composition. The composition can be used as a
XX CC medicament (or in the manufacture of a medicament) for treating,
XX CC preventing or diagnosing infection due to Neisserial bacteria. For
XX CC example, some of the identified proteins could be components of vaccines
XX CC against Meningococcus B; against all serotypes; and/or against all
XX CC pathogenic Neisseriae. Identification of sequences from the bacterium
XX CC will also facilitate production of biological probes, particularly
XX CC organism-specific probes. Attempts to make efficacious Meningococcus B
XX CC vaccines have failed mainly due to antigen tolerance. Multivalent
XX CC vaccines have also been tried but none have successfully overcome
XX CC antigenic variability. The provision of further, complete sequences may
XX CC provide an opportunity to identify secreted or surface exposed proteins
XX CC that may be presumed targets for the immune system and which are not
XX CC antigenically variable or at least more conserved than other more
XX CC variable regions
XX SQ Sequence 35133 BP; 8216 A; 9873 C; 8987 G; 8057 T; 0 U; 0 Other;
Query Match 76.3%; Score 20.6; DB 3; Length 35133;
Best Local Similarity 85.2%; Pred. No. 14;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 GTCGAACGAGACTTCCCAACCGAC 27
Db 30997 GTCGAACGAGACTTCCCAACCGAC 31023
RESULT 9
AAA81489.3
Continuation (4 of 9) of AAA81489 from base 300001 (N. meningitidis partial DNA sequence

WP	Sequence split into 9 fragments	LOCUS AAA81489	Accession	AAA81489
WP	Fragment Name	Begin	End	
WP	AAA81489_0	1	110000	
WP	AAA81489_1	100001	210000	
WP	AAA81489_2	200001	310000	
WP	AAA81489_3	300001	410000	
WP	AAA81489_4	400001	510000	
WP	AAA81489_5	500001	610000	
WP	AAA81489_6	600001	710000	
WP	AAA81489_7	700001	810000	
WP	AAA81489_8	800001	837096	

```
Query Match      76.3%; Score 20.6; DB 3; Length 110000;
Best Local Similarity 85.2%; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 4; Indels 0;
```

QY 1 GTCGAACGAGACTTCCCAAAACCGAC 27
 ||| ||||| ||||| ||||| |||||
Db 49086 GTCAAACGAGACTTGCCGAAGCCGAC 49112

RESULT 10
AAAF21611
ID AAAF21611 standard; DNA; 349980 BP.

CC and/or antibodies which binds to the proteins can be used in compositions
CC for treating or preventing infection due to Neisserial bacteria or as a
CC diagnostic reagent for detecting the presence of Neisserial bacteria or
CC of antibodies raised to Neisserial bacteria. Computers, computer memory,
CC computer storage medium or computer databases can be used in a search to
CC identify open reading frames (ORFs) or coding sequences within the NMB
CC genome. The DNA sequences provide further opportunities to find antigenic
CC or immunogenic proteins which are more effective in vaccines than the
CC outer membrane proteins currently used
XX
SQ Sequence 349980 BP; 87189 A; 93501 C; 84627 G; 84663 T; 0 U; 0 Other;

```
Query Match      76.3%; Score 20.6; DB 3; Length 349980;
Best Local Similarity 85.2%; Pred. No. 19;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0
```

Qy 1 GTCGAACGAGACTTTCCCAACCGAC 27
||| ||||| ||||| ||||| |||||
Db 284315 GTC AACGAGACTT TGC CGAAGCCGAC 284341

RESULT 11
AAS46299/c
ID AAS46299 standard; DNA; 6862 BP.

RESULT 11
AAS46299/c
ID AAS46299 standard; DNA: 6862

CC diseases or the predisposition to specific diseases, by analysing
CC cytosine methylations. The parameters may be compared to another set of
CC genetic and/or epigenetic parameters, the differences serving as basis
CC for diagnosis and/or prognosis events which are disadvantageous to
CC patients. The present sequence is one of the 533 genomic sequences
CC derived from tumour suppressor genes and oncogenes. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6862 BP; 1278 A; 518 C; 2176 G; 2890 T; 0 U; 0 Other;

Query Match 74.8%; Score 20.2; DB 4; Length 6862;
Best Local Similarity 88.0%; Pred. No. 16;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAAGGAGACTTCCCAACCGA 26
DB 6351 TCGAAGGAGCTTTACCAACCGA 6327

RESULT 12
ABL32222/c
ID ABL32222 standard; DNA; 6862 BP.
XX
AC ABL32222;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 195.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antartiosclerotic; antihaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 195; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 6862 BP; 1278 A; 518 C; 2176 G; 2890 T; 0 U; 0 Other;

Query Match 74.8%; Score 20.2; DB 6; Length 6862;
Best Local Similarity 88.0%; Pred. No. 16;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAAGGAGACTTCCCAACCGA 26
DB 6351 TCGAAGGAGCTTTACCAACCGA 6327

RESULT 13
AAS61081/c
ID AAS61081 standard; DNA; 6862 BP.
XX
AC AAS61081;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #36.
XX

KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma, HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Freckleplasia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.

PN WO200177375-A2.

PD 18-OCT-2001.

PF 06-APR-2001; 2001WO-EP003968.

PR 06-APR-2000; 2000DE-01019058.

PR 07-APR-2000; 2000DE-01019173.

PR 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-017470/02.

PT New nucleic acid sequences from chemically modified genes associated with
PT gene regulation, useful for analyzing cytosine methylations for diagnosis
PT and therapy of diseases e.g. severe combined immunodeficiency disease.

PS Claim 1; SEQ ID NO 37; 26pp; English.

CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The chemical
CC pretreatment converts cytosine bases unmethylated at the 5-position to
CC uracil or another base with hybridisation behaviour dissimilar to
CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
CC oligomers (or sets/arrays) and method are useful in the diagnosis of
CC diseases (or predisposition to diseases) associated with gene regulation
CC and in therapy of such diseases, by enabling analysis of the cytosine
CC methylation patterns of such genes, kits are provided. They are
CC especially useful in diagnosis and therapy of e.g. severe combined
CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
CC syndrome, renal disease, pre-eclampsia, graft versus-host disease. The
CC present sequence is a sequence included in the sequence data for this
CC specification and is associated with the human gene regulation-associated
CC genes. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX

Matches	21;	Conservative	0;	Mismatches	4;	Indels	0;	Gaps	0;
Qy	2	TCGAACGAGACTTTCCGCCAAACCGA	26						
Db	6611	TCGAATGAGAAATTTCCGCCAAGCCCA	6635						

Search completed: April 29, 2004, 04:56:49
 Job time : 226.982 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:25:04 ; Search time 407.435 Seconds
(without alignments)
2872.264 Million cell updates/sec

Title: US-10-624-714-1
Perfect score: 27
Sequence: 1 gtcgaacgagactttccccaaccgac 27

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hgt: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_ov: *
22: em_or: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rod: *
36: em_htg_man: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	
1	27	100.0	16644	1	AE006918 Mycobacte	
C	27	100.0	38721	1	MSGY219	
	27	100.0	341957	15	BX842572 Mycobacte	
	27	100.0	343050	1	BX248334 Mycobacte	
5	20.6	76.3	6972	1	NMU58765 Neisseria m	
6	20.6	76.3	14056	1	AE002521 Neisseria	
7	20.6	76.3	326301	1	NMA522491 Neisseria	
8	20.6	76.3	349980	6	AX044033 Sequence	
C	20.2	74.8	6862	6	AX251053 Sequence	
C	20.2	74.8	6862	6	AX251776 Sequence	
C	11	20.2	74.8	6862	6	AX345124 Sequence
C	12	19.2	71.1	179886	2	AC140072 Mus muscu
C	13	19.2	71.1	219568	10	AL844840 Mouse DNA
C	14	19	70.4	1382	1	TNBLA
C	15	19	70.4	3512	8	AK110440 Oryza sat
C	16	19	70.4	4846	1	AF039487 Thermotoga
C	17	19	70.4	308399	2	AC110098 Rattus no
C	18	19	70.4	311600	1	AE016871 Pseudomon
C	19	18.8	69.6	71352	8	NCBI10C3 Neutropor
C	20	18.8	69.6	247582	2	AC115339 Rattus no
C	21	18.6	68.9	60	9	AF064918 Homo sapi
C	22	18.6	68.9	1002	8	AF249273 Helianthu
C	23	18.6	68.9	1086	8	AJ250224 Helianthu
C	24	18.6	68.9	2998	5	AB120744 Carassius
C	25	18.6	68.9	3173	8	AB196454 Laminaria
C	26	18.6	68.9	46275	2	AC014969 Drosophil
C	27	18.6	68.9	163709	3	AC108877 Drosophil
C	28	18.6	68.9	187437	3	AC023751 Drosophil
C	29	18.6	68.9	22922	2	AC106497 Rattus no
C	30	18.6	68.9	302417	3	AE003482 Drosophil
C	31	18.2	67.4	3019	8	BT005794 Arabidops
C	32	18.2	67.4	3329	8	AY080791 Arabidops
C	33	18.2	67.4	3348	8	AY050658 Arabidops
C	34	18.2	67.4	123943	10	AL592547 Mouse DNA
C	35	18.2	67.4	124550	8	AC134885 Genomic s
C	36	18.2	67.4	135611	14	AF169823 Spodopter
C	37	18.2	67.4	138181	8	AC005171 Arabidops
C	38	18.2	67.4	178967	2	AC135864 Oryza sat
C	39	18.2	67.4	190870	2	AC128850 Rattus no
C	40	18.2	67.4	198050	1	AL646061 Ralstonia
C	41	18.2	67.4	206563	2	AC106494 Rattus no
C	42	18.2	67.4	271208	2	AC126847 Rattus no
C	43	18	66.7	1247	8	JVSSSRNAG1 Juniperus
C	44	18	66.7	2319	8	AB029368 Juniperus
C	45	18	66.7	2725	8	AB029366 Chamaecypr

ALIGNMENTS

RESULT 1
AE006918
LOCUS
DEFINITION
Mycobacterium tuberculosis CDC1551, linear BCT 27-APR-2001
16644 bp DNA section 4 of 280 of the
complete genome.
ACCESSION
AE006918 AE000516
VERSION
AE006918.1 GI:13879090
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis CDC1551
Mycobacterium tuberculosis CDC1551
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1 (bases 1 to 16644)
REFERENCE
AUTHORS
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,

Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Unayam, L.A., Emolaeva, M., Khouri, H., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
Unpublished
2 (bases 1 to 16644)
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Unayam, L.A., Emolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1. 16644
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/notes="clinical strain"
complement (86..712)
/gene="MT0048"
/genes="MT0048"
/notes="similar to PTD:2440102; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, Marr family"
/protein_id="AAK44270.1"
/db_xref="GI:13879091"
/translation="MSVRSIGKMKRISGNALAVKGRPTQVYGHVRLDGRFAD SEPTAEVQLAELHRALEKILMLRGDPNGAGDLTLAQLSLVTLVLLDGPTRM TDIAHVRNPTTVAIRLEKILGVKSRSFSDRAVLVITPQGRAVHESLANR RAALALLQSLPSDLTUKALAPLERLASGEPASGASPNAPARRA"
complement (871..1605)
/gene="MT0049"
complement (871..1605)
/gene="MT0049"
/notes="similar to GP:2623054; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, GntR family"
/protein_id="AAK44271.1"
/db_xref="GI:13879092"
/translation="MPKKYGVKEQDVVAHILNLLTGKLRSGDRVDRNEIAGLGVS RVPIQALVLEHDGIVSTRYHGRFAIEPFDVATILEHDELGLNGIASAAANPT PRILQDVAWLSRKSERAPAEVCWEYRRTVNDYAGPRLHATIRASQNLIPRV WMYQSRDDVLPFPYEENAAIHRREPEAARACIGRSELMAQTMLAELEFRRLVLPV EGACPGPGFAPIGFARSQPSPPV"
complement (1738..2532)
/gene="MT0050"
complement (1738..2532)
/gene="MT0050"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44272.1"
/db_xref="GI:13879093"
/translation="MTSLVRDPLVPIVQLQPOQAPHYRAVDVRCEDIGVDIAF TDHFFPLVDPDPGPFECWTVLGAWEQTSHEICALVTCNSYNPELLAMATVD HSGGRLLIGSGWKQDYGYRFGTAGSLRDLAALPRIKARLGNPPTRD IPVLIGGGERKLRIVAYADIWHSFTAGSDYLAKSAVLSTHCSTVGRNPATIBRSA AVDGGGLIASAEALAGLVTLITVCGDGDYDLSAAALCRWRDRG"
complement (2548..3444)
/gene="MT0051"
complement (2548..3444)
/gene="MT0051"

/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44273.1"
/db_xref="GI:13879094"
/translation="MLSDDELTDGDEFALLAENAEQAGVNGPLPEVERVOAGATSAUR WGSAPRVIFLHGGQNAHTWTVIVGLGPALAVDLPGHSHSARREDGNYSPOLNSE TLAPVILRELACAEFFVYVMSLGLTAIRLAAMPDLVGLVLDVTSALQRAELTA EQRTVALMHGEREFPSQAWLDTIAAHPDRDVKLSRRGVFNHNRDLNNGNWRVYD AIRTFGDFAGLWDDVDAISAPITLVROGSGSFVTDQDTABLHRRATHFRGWHVKSQ HSVDQDPRALIEIVRGVLDTR"
complement (3526..4629)
/gene="MT0052"
complement (3526..4629)
/gene="MT0052"
/note="similar to GP:4808395; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="1L-myo-inositol-1-phosphate synthase"
/protein_id="AAK44274.1"
/db_xref="GI:13879095"
/translation="MSEHQSLPAPASTEVRVAIVGVNCASSLIVQGVVEYYNADDTSTVPGLMHVRFGPYHVRDVKFAAFDPAKVPDLSDAIFASENTIKIADVAPTNI VQRTPLDGIKGYADTIELSDAPVDVQALKEAKVDVLSVYLPVGESEADRFYQOC AIDAGVAFVNALPVFIASDPWAKKFTDAGVPIVGDIDKQVGATITHRVLAKLFEDR VQIJDRTMQLNVGNGMDFLANLEERLESKKIKTQAVTSNLRKPEKTKDVHIGPSDH VGMIDRKMAYVLEEGAFGDVPLNLEYKLEVDSPNSAGVIIDAVPAKAIKDRGIG GPVTPASAYLMKSPFQPLDDIARAQLEEFII"
complement (4690..5232)
/gene="MT0053"
complement (4690..5232)
/gene="MT0053"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44275.1"
/db_xref="GI:13879096"
/translation="MLELAILGLLIESPMHGYELRKRLTGVLGAFRAFSYGLYPALR RMQADGLIAENAPAGTPVRRARVYOLTDKRRRFGELVADTGPHNYTDDGFGVHLA FNNTPABARMLEIGRRRQVRERREGLEAVARASSFDTRYTRQLHQLGLESSEREV KWLNELIAERAPNPAPAEQT"
complement (5333..6202)
/gene="MT0054"
complement (5333..6202)
/gene="MT0054"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44276.1"
/db_xref="GI:13879097"
/translation="MAKWIGAPLARGVSTATRAKDSRODACHILDALRDGELSMEE HRSEVSATKAVTLGDLQRLVADLOVESAPQWPKRAKTELGLIAAFVAVSLV HGIGKGVYNTASPLDFTDPGAKPDGIAPVLTTPRQLHSLGGITGLLEQTRKRFQ DTMGRUVIIPYASLDVRDPADRVLTATYRGGWDATSSAKSADVSVDLSKFD AKTAVGIMRGAPETLGLKQSDVKSMYLIVDFAKDPTTFAALSLSLVSSDYGGGYLVF AGDTIKHVSYPs"
6336..6749
/gene="MT0055"
6336..6749
/gene="MT0055"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44277.1"
/db_xref="GI:13879098"
/translation="MDYTLRRSILAAVYSGRTGVSECDANPYLLAAKFEKPKPSRV ICPIKKEQUTLVSWFGHGLVAGSANTAEELILLATRFSEFAHVHVEVCRTCSWN HLKVSYVLGAARPARPPRGSGGTRTARNGARTASE"


```
gene      6742..9204
          /gene="MT0056"
CDS       6742..9204
          /gene="MT0056"
          /note="similar to PID:886307; identified by sequence
          similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="penicillin-binding protein"
          /protein_id="AAK44278.1"
          /db_xref="GI:13879099"
          /translation="MNSDGRHQSSGAPRGANFGQGVPPDRRLTAIPLPVTDDR
          SAPHASTBAVAALDGPMPFPDPLEEVTAALAAPFGPPRGDQLGRRRPPGP
          GPFGSQPAGLPRVDFPRVQINMKIRSLYLAAVVILLPMVFTFYALIVD
          VPKGDIRNQVSTIIASDSEWAKIVPEEGNEVDNLQSVPMHYRQAVIAEDRNFY
          SNPGFSFTAFANKNLIPFGDGGSTITQYVKNALVGSQHGWSGLMKAKELVI
          ATKMSGWSKDLQAYLNIIFYGRGAVGISAAKAYPKPVEQLTVAGALLAALIR
          RSTLDPADVPGAHARNWLDGMVETKALSPNDRAAQVPEPTVPDLAENCTKG
          PNLGLERQIRLELLEFNIDEQLTNQGLVWTTIDPQORAAEKAVAKYLDGQPDPM
          RAAVSDIPHNAGRAYEGDNANGFDFAQAGLQTSQSFVKFALVALEQGIQYQV
          DSSPLTVDGKITNVEGECGTCNTAALKKMSLNTSYRLMLKINGSPQAVADAHQ
          GIASPPGVGVAHTLSDGKGPNGIVLQYQYTRVIDMASAYATLAASGIYHPHFVQ
          KVSANGQVLPDASTADNGTDQRIKPAVDNVTAAEPIAGYSRGNLAGGRDSAAKT
          GTQFGDITANKDAMVGTPTSLTAVWGTGKDEPLVTASGAALVSGSLPSDIWKA
          TWDGALKGTSNETFPKPTBVGAGVPPPPPEPSEVPTVLTQVTEIAPGITPIGP
          PTTITLAPPPPPAPPAATPTPP"
          9246..10883
          /gene="MT0057"
          9246..10883
          /gene="MT0057"
          /note="identified by Glimmer2; putative"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AAK44279.1"
          /db_xref="GI:13879100"
          /translation="VAADLRSDNRDPSRTDVLGAALANVVGVGPRHALIGTRLM
          TELRVFALFLALGWSKACLOSTGPGDORVANWDRAYVOLCYSDTVPLV
          GAELLSGQFPKYSWETSDNGTQPLRDQGLAVRYMEYVLTGIYQLVSNIAKTY
          TALSKGFLVAEVVFPNVAAGLAWLITVWATSGLAGRWIDALVAASPLVI
          FOITNFADLATGATSGLLAWRRPVLAVGLIGLSAAKLYPLFLLELLGIRA

Query Match      100.0%; Score 27; DB 1; Length 16644;
Best Local Similarity 100.0%; Pred. No. 0.045;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1  GTCGAACGAGACTTCCCAAAACCGAC 27
      |||||
Db      8988 GTCGAACGAGACTTCCCAAAACCGAC 9014

RESULT 2
MSGY219/c      38721 bp      DNA      linear      BCT 10-DEC-1996
LOCUS      Mycobacterium tuberculosis sequence from clone y219.
DEFINITION      AD000013
ACCESSION      AD000013.1  GI:1717736
KEYWORDS      Mycobacterium tuberculosis
SOURCE      Mycobacterium tuberculosis
ORGANISM      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
      Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
      tuberculosis complex.
      1 (bases 1 to 38721)
REFERENCE      Du,L.
AUTHORS      Direct Submission
TITLE      Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
JOURNAL      Beaver Street, Waltham, MA, USA, 02154 du@cric.com
COMMENT      GSDB:S:1004710
FEATURES      Location/Qualifiers
      source      1..38721
      /organism="Mycobacterium tuberculosis"
      /mol_type="genomic DNA"

/db_xref="taxon:1773"
/clone="y219"

Query Match      100.0%; Score 27; DB 1; Length 38721;
Best Local Similarity 100.0%; Pred. No. 0.049;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ORIGIN
QY      1  GTCGAACGAGACTTCCCAAAACCGAC 27
      |||||
Db      6240 GTCGAACGAGACTTCCCAAAACCGAC 6214

RESULT 3
BX842572
ID      BX842572      standard; circular genomic DNA; PRO; 341957 BP.
XX      BX842572; AL021427; AL021428; AL021926; AL021927; AL021928; AL021929;
AC      AL021930; AL223456; Z74410; Z80233; Z80775; Z86089; Z92669; Z92671;
AC      Z97050;
XX      BX842572.1
XX      21-NOV-2003 (Rel. 77, Created)
DT      21-NOV-2003 (Rel. 77, Last updated, Version 1)
DT      21-NOV-2003 (Rel. 77, Last updated, Version 1)
XX      Mycobacterium tuberculosis H37Rv complete genome; segment 1/13
DE      complete genome.
XX      complete genome.
XX      Mycobacterium tuberculosis H37Rv
OS      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC      Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC      Mycobacterium tuberculosis complex; Mycobacterium tuberculosis.
XX      [1]
RX      MEDLINE; 98295987.
RX      PUBMED; 9634230.
RA      Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA      Gordon S.V., Eigmeier K., Gas S., Barry III C.E., Tekala F., Badcock K.,
RA      Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K.,
RA      Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K.,
RA      Krogh A., McLean J., Moule S., Murphy J., Oliver S., Osborne J.,
RA      Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S.,
RA      Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA      "Deciphering the biology of Mycobacterium tuberculosis from the complete
RT      genome sequence";
RT      Nature 393:537-544(1998).
XX      [2]
RX      PUBMED; 12368430.
RX      Camus J.C., Pryor M.J., Medigue C., Cole S.T.;
RA      "Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv";
RT      Microbiology 148:2967-2973(2002).
XX      [3]
RP      1-341957
RA      Parkhill J.;
RT      Submitted (11-JUN-1998) to the EMBL/GenBank/DBJ databases.
RL      Submitted on behalf of the Mycobacterium tuberculosis sequencing and
RL      mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton,
RL      Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut
RL      Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail:
RL      parkhill@sanger.ac.uk
XX      Notes:
CC      Details of M. tuberculosis sequencing at the Sanger Centre
CC      are available on the World Wide Web.
CC      (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/)
XX      Key      Location/Qualifiers
FH      Key      Location/Qualifiers
```

FT source 1. .341957
 FT /db_xref="taxon:83332"
 FT /mol_type="genomic DNA"
 FT /organism="Mycobacterium tuberculosis H37Rv"
 FT /strain="H37Rv"
 FT CDS 1. .1524
 FT /evidence=EXPERIMENTAL
 FT /note="RV0001, MT0029.01, P49993, len: 507 aa.
 FT dnaA, chromosomal replication initiator protein (see
 FT citations below), equivalent to other Mycobacterial
 FT CHROMOSOMAL REPLICATION INITIATOR PROTEINS e.g.
 FT P46388 [DNAA MYCLE from Mycobacterium leprae (502 aa);
 FT Q9U17L [DNAA MYCFA from Mycobacterium paratuberculosis (509
 FT aa); P49990 [DNAA MYCAV from Mycobacterium avium (508 aa);
 FT P49992 [DNAA MYCSM from Mycobacterium smegmatis (504 aa);
 FT etc. Also highly similar to others except in N-terminus
 FT e.g. Q9ZH75 [DNAA STRCH CHROMOSOMAL REPLICATION INITIATOR
 FT PROTEIN from Streptomyces chrysomallus (624 aa);
 FT Q9ZH76 [DNAA STRE from Streptomyces reticuli (643 aa);
 FT DNAA_ECOLI [P03004|B3702 chromosomal replication initiator
 FT protein from Escherichia coli strain K12 (467 aa), FASTA
 FT scores: opt: 986, E(): 0, (43.2% identity in 389 aa
 FT overlap); etc. Contains PS00017 ATP/GTP-binding site motif
 FT A (P-loop) and PS01008 DnaA protein signature. BELONGS TO
 FT THE DNAA FAMILY. Note that the first base of this gene has
 FT been taken as base 1 of the Mycobacterium tuberculosis
 FT H37Rv genomic sequence."
 FT /transl_table=11
 FT /function="PLAYS AN IMPORTANT ROLE IN THE INITIATION AND
 FT REGULATION OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN
 FT OF REPLICATION; IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA
 FT AT A 9 BP CONSENSUS (DNAA BOX): 5'-TTATC(C/A)(C/A)A-3'.
 FT DNAA BINDS TO ATP AND TO ACIDIC PHOSPHOLIPIDS. DNAA PROTEIN
 FT BINDS THE ORIGIN OF REPLICATION (oriC), ATP AND ADP, AND
 FT EXHIBITED WEAK ATPASE ACTIVITY."
 FT /gene="dnaA"
 FT /locus_tag="Rv0001"
 FT /product="CHROMOSOMAL REPLICATION INITIATOR PROTEIN DNAA"
 FT /protein_id="CAA16238.1"
 FT /translation="MTDDPGSGFTVWNAVSELNPKVDDPQSDANLSAPLPQOR
 FT AWNLVQPLTVIGFALLSPVSSVQNEIHLRAPITDALSRRGHQIQGVRTAPPA
 FT TDEADDTVPSPNPATTPDPTTNDDEIDDSAAAGDNOHSPSYFTRPHNTDSATA
 FT GVTSLNRTFTDFTVIGANRFAHAALAAIAEAPARAYNPLFWGESGLGKTHLHAAG
 FT NYAQLFPGRVYVTEETNDPINSLDRKVAFKRSYRDVLLVDVDFIQIFEGREG
 FT IDEFHTFTLNANKQIVISDRPPKQATLEDLRTFWEGLITDPVQPELETRIA
 FT ILKQAKMERLAVPDVLLVLIASSIRNIRELEGALIRVTAFASLTKPTIDKALAEIVL
 FT RLIIADANMTOISAAITMAATAYFTTVEELRGPKTKALAQSIAMYLCRELTDL
 FT LPKIGAFGRDHTTMYAQKILSEMAERREVDHVKELTTIRQSKR"
 FT 622..645
 FT /notes="PS00017 ATP/GTP-binding site motif A (P-loop)"
 FT 1384..1440
 FT /notes="PS01008 DnaA protein signature"
 FT 2052..3260
 FT /evidence=EXPERIMENTAL
 FT /note="RV0002, (MTV029.02, MTCY10H4.0), len: 402 aa. DnaA.
 FT DnaA polymerase III (beta chain) (EC 2.7.7.7) (see citations
 FT below), equivalent to other Mycobacterial DNA POLYMERASES
 FT III BETA CHAIN e.g. NP_301130.1|NC_002677 from
 FT Mycobacterium leprae (399 aa); Q9L7L6|DP3B MYCPA from
 FT Mycobacterium avium subsp. paratuberculosis (399 aa);
 FT P52851|DP3B MYCSM from Mycobacterium smegmatis (397 aa);
 FT etc. Also highly similar to others e.g. P27903|p3B STRCO
 FT DNA POLYMERASE III BETA CHAIN from Streptomyces coelicolor
 FT (376 aa), FASTA scores: opt: 1189, E(): 0, (52.8% identity
 FT in 337 aa overlap); P21174|p3B Miciu from Micrococcus
 FT luteus (310 aa); P52023|p3B SYNPF from Synecococcus sp.
 FT strain PCC 7942 (375 aa); etc. Overlaps and extends CDS in
 FT neighbouring cosmid MTCY10H4.01."
 FT /transl_table=11
 FT /EC_number="2.7.7.7"
 FT /function="DNA POLYMERASE III IS A COMPLEX, MULTICHAIN
 FT ENZYME RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN
 FT BACTERIA. THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5'

FT EXONUCLEASE ACTIVITY. THE BETA CHAIN IS REQUIRED FOR
 FT INITIATION OF REPLICATION ONCE IT IS CLAMPED ONTO DNA, IT
 FT SLIDES FREELY (BIDIRECTIONAL AND ATP-INDEPENDENT) ALONG
 FT DUPLEX DNA [CATALYTIC ACTIVITY: N deoxynucleoside
 FT triphosphate = N diphosphate + {DNA}N]."
 FT /gene="dnaA"
 FT /locus_tag="Rv0002"
 FT /product="DNA POLYMERASE III (BETA CHAIN) DNAN (DNA
 FT NUCLEOTIDYLTRANSFERASE)"
 FT /protein_id="CAA16239.1"
 FT /translation="MDAATTRVGLTDTLFRLLRESFADAVSWAKNLPARPAPVPLSGV
 FT LLTGSDNGLITSGFDEVSAAQVABISVPSGLVSLGRLSDITRALPNKPDVHVVEG
 FT NRVALTCGNARFVSLTPMPVEDYPTLPTLPEETGLPAELFAEISVAIAAGRDDTLPM
 FT LTRGVILGETVTLVLAATDFRLAVLKWSSASPDIEARVLPKATLAEAAKAGIGGS
 FT DVRLSLGTGPGVKGGLGSGKSGSTTRLLDRERPKFQLLPTEHTVATVMDVAELI
 FT EAIKLVALVADRAQVRVEFADSGVLSAGADDVGRAEEDLVVDYAGEPTTIAFNPTYL
 FT TDGLSSRSRVSFGFTTAGKALLRPVSGDDRPVAGLNGNGPFPVSTVDYVYLLMPVR
 FT LPG"
 FT 3280..4437
 FT /evidence=EXPERIMENTAL
 FT /note="RV0003, (MTCY10H4.01), len: 385 aa. recF, DNA
 FT replication and repair protein (see citations below),
 FT equivalent to others Mycobacterial DNA replication and
 FT repair proteins e.g. NP_301131.1|NC_002677 from
 FT Mycobacterium leprae (385 aa); Q9L7L5|RECF MYCPA from
 FT Mycobacterium avium subsp. paratuberculosis (385 aa);
 FT P50916|RECF MYCSM from Mycobacterium smegmatis (384 aa);
 FT etc. Also highly similar to others e.g. P36176|RECF_STRCO
 FT DNA REPLICATION AND REPAIR PROTEIN from Streptomyces
 FT coelicolor (373 aa); NP_440892.1|NC_000911 from
 FT Synechocystis sp. strain PCC 6803 (384 aa);
 FT NP_469352.1|NC_003212 from Listeria innocua (370 aa); etc.
 FT Contains PS00017 ATP/GTP-binding site motif A (P-loop),
 FT PS00617 RecF protein signature 1, and PS00618 RecF protein
 FT signature 2. BELONGS TO THE RECF FAMILY."
 FT /transl_table=11
 FT /function="THE RECF PROTEIN IS INVOLVED IN DNA METABOLISM
 FT AND RECOMBINATION; IT IS REQUIRED FOR DNA REPLICATION AND
 FT NORMAL SOS INDUCIBILITY. RECF BINDS PREFERENTIALLY TO
 FT SINGLE-STRANDED, LINEAR DNA. IT ALSO SEEMS TO BIND ATP."
 FT /gene="recF"
 FT /locus_tag="Rv0003"
 FT /product="DNA REPLICATION AND REPAIR PROTEIN RECF
 FT (SINGLE-STRAND DNA BINDING PROTEIN)"
 FT /protein_id="CA302424.1"
 FT /translation="MYVRHGLRDFRSWACVDLEHLEHGRTVFVPGNGYKTNLEALWY
 FT STTLGHRVSGADPLIRVGTDRAVITIVVNDGRCVADLEIATGRVKNRLNRSSVRS
 FT TRDVGVHRAVLPAEDLGLVRGDPADRRVYLDLAIVRPATAVAEVRVLRQRTA
 FT LLKSVGARYRGDRGVPTLEWVDSRLAEHGAELVAARDLVNQLAPEVKKAYQLLAPZ
 FT SRASGIGYRASMDVTPGSEOSIDROLLAARLAAALAAARDAELEGVCLVGGPHRODLI
 FT LRLGDDPAKGFASHGEAWSLAVALAAVQLRVDGGEVPLLDLDDFAELDVNRRALLA
 FT TAAASEQVLVTAAVLIEDIPAGWDARVRHIDVRADDTGSSVVLP"
 FT 3367..3390
 FT /notes="PS00017 ATP/GTP-binding site motif A"
 FT 3634..3690
 FT /note="PS00617 RecF protein signature 1"
 FT 4243..4296
 FT /note="PS00618 RecF protein signature 2"
 FT 4434..4997
 FT /evidence=EXPERIMENTAL
 FT /note="RV0004, (MTCY10H4.02), len: 187 aa. Conserved
 FT hypothetical protein (see Salazar et al., 1996), highly
 FT similar, but longer 21 aa in N-terminus, to
 FT AAF3896.1|AF222789 unknown protein from Mycobacterium
 FT avium subsp. paratuberculosis (166 aa); and highly similar
 FT to NP_301132.1|NC_002677 conserved hypothetical protein
 FT from Mycobacterium leprae (189 aa); S70990 hypothetical
 FT protein from Mycobacterium smegmatis (194 aa). Also highly
 FT similar, except in N-terminal part, to

Query Match 100.0%; Score 27; DB 15; Length 341957;
 Best Local Similarity 100.0%; Pred. No. 0.063;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGAACGAGACTTTCCCAACCGAC 27
 |||||
 Db 55483 GTCGAACGAGACTTTCCCAACCGAC 55509

RESULT 4
 BX248334 343050 bp DNA linear BCT 11-JUN-2003
 LOCUS Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment
 DEFINITION 1/14.

ACCESSION BX248334 BX248333
 VERSION BX248334.1 GI:31616762
 KEYWORDS complete genome.
 SOURCE Mycobacterium bovis subsp. bovis AF2122/97
 ORGANISM Mycobacterium bovis subsp. bovis AF2122/97
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
 tuberculosis complex.

REFERENCE 1
 Garnier,T., Eglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
 Pryor,M., Duchoy,S., Grondin,S., Iacroix,C., Monsempe,C., Simon,S.,
 Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,
 Wheeler,P.R., Parkhill,J., Barrell,B.G., Cole,S.T., Gordon,S.V. and
 Hewinson,G.
 The complete genome sequence of Mycobacterium bovis
 Online Publication
 PNAS 10.1073/pnas.1130426100 (Microbiology)
 2 (bases 1 to 343050)

TITLE Direct Submission
 JOURNAL Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
 Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex
 15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
 Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
 Laboratories Agency Weybridge, Woodham Lane, New Haw,Addlestone,
 Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
 Hinxton,Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut
 Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
 Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
 rue du Docteur Roux, 75724 Paris Cedex 15, France
 Location/Qualifiers

FEATURES
 source
 1. 343050
 /organism="Mycobacterium bovis subsp. bovis AF2122/97"
 /mol_type="genomic DNA"
 /strain="AF2122/97"
 /db_xref="taxon:233413"
 1. 1524
 /gene="dnaA"
 /locus_tag="Mb00001"
 1. 1524
 /gene="dnaA"
 /locus_tag="Mb00001"
 /note="Mb0001. dnaA, len: 507 aa. Equivalent to Rv0001,
 len: 507 aa, from Mycobacterium tuberculosis strain H37Rv,
 (99.6% identity in 507 aa overlap). dnaA, chromosomal
 replication initiator protein (see citations below),
 equivalent to other Mycobacterial CHROMOSOMAL REPLICATION
 INITIATOR PROTEINS e.g. P46388|DnaA MYCFA from
 Mycobacterium leprae (502 aa); Q9L7L7|DnaA MYCRA from
 Mycobacterium paratuberculosis (509 aa); P49990|DnaA MYCAV
 from Mycobacterium avium (508 aa); P49992|DnaA MYCSM from
 Mycobacterium smegmatis (504 aa); etc. Also highly similar
 to others except in N-terminus e.g. Q9ZH75|DnaA_STRCH
 CHROMOSOMAL REPLICATION INITIATOR PROTEIN from
 Streptomyces chrysomallus (624 aa); DnaA ECOLI|P03004|B3702
 Streptomyces reticuli (643 aa); DnaA ECOLI|P03004|B3702
 chromosomal replication initiator protein from Escherichia
 coli strain K12 (467 aa), FASTA scores: opt: 986, E(): 0,
 (43.2% identity in 389 aa overlap); etc. Contains PS00017
 ATP/GTP-binding site motif A (P-loop) and PS01008 DnaA
 protein signature. BELONGS TO THE DnaA FAMILY. Note that
 the first base of this gene has been taken as base 1 of

the Mycobacterium bovis genomic sequence."

/codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="CHROMOSOMAL REPLICATION INITIATOR PROTEIN DnaA"
 /protein_id="CAD92863.1"
 /db_xref="GI:31616763"
 /translation="MTDPSGFTTVMNVAVSELNGPKVDGDPSSDANLSAPLTPOQ
 RANLNLPQTLIVEGFALLSVPSFVQNEIERHLRAIPITDALSRLGHQIQGVRIAP
 PATADADITVPSENPAITSPOTTINDDEIDUSAAAGNGNQHSWPSYFTTERPENTDS
 ATAGVSLNRRYTFDTFVIGASNRFAHAALAAEAPARVNPFLIFGESGLKTHLL
 HAAGYARLFFPGMRVYKVTSEETNFINSRDDRVAFKFSRVDDVLVDDIQT
 EGKEGIQREFFHTNHNANKQIVISDRPPKQLATLEDRLETRFEWGLTKTPDIK
 LETRIALRKQAOMERLAI PDDVLELIASIERINRELEGALIRVTAFASTNKTPIDK
 ALAEIVRLDIADANTQIISAATIMAATAEVDITVEELRGPGRKTRALASQRIAMYL
 CRELTDLSLPAIGQAFORDHTTVMYQAKLISEMAERREVDYHVELTTRIQSSKR"
 2052. 3260
 /gene="dnaN"
 /locus_tag="Mb00002"
 2052. 3260
 /gene="dnaN"
 /locus_tag="Mb00002"
 /EC_numbers="2.7.7.7"
 /note="Mb00002, dnaN, len: 402 aa. Equivalent to Rv0002,
 len: 402 aa, from Mycobacterium tuberculosis strain H37Rv,
 (99.8% identity in 402 aa overlap). dnaN, DNA polymerase
 III (beta chain) (EC 2.7.7.7) (see citations below),
 equivalent to other Mycobacterial DNA POLYMERASES III BETA
 CHAIN e.g. NP_301130.1|NC_002677 from Mycobacterium leprae
 (399 aa); Q9L7L6|DP3B MYCFA from Mycobacterium avium
 subsp. paratuberculosis (399 aa); P52851|DP3B MYCSM from
 Mycobacterium smegmatis (397 aa); etc. Also highly similar
 to others e.g. P27903|DP3B_STRCO DNA POLYMERASE III BETA
 CHAIN from Streptomyces coelicolor (376 aa), FASTA scores:
 opt: 1189, E(): 0, (52.8% identity in 337 aa overlap);
 P21174|DP3B MICLU from Micrococcus luteus (310 aa);
 P52023|DP3B_SINP7 from Synecococcus sp. strain PCC 7942
 (375 aa); etc. Overlaps and extends CDS in neighbouring
 cosmid MTCY10H4.01."
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="DNA POLYMERASE III (BETA CHAIN) DnaN (DNA
 NUCLEOTIDYLTRANSFERASE)"
 /protein_id="CAD92864.1"
 /db_xref="GI:31616764"
 /translation="MDAATRVGLTDLTFLRLRESFADAVSWAKNLPARPVPLSG
 VLLTGSNDGLTISGFDYEVSAEAOVGAIVSGRLSDITRALPNKPVGVHV
 EGNRVATCGNARSLPTMPVEDYPTLPTPEETGLLPDAEIAISOVAALAGRDDT
 LPMLTGIVEILGETVILVLAADRFLAVRELKWSASSPDIEAIVLPKTLAEAKAG
 IGGSDVRUSLGTGPGVKDGLUGISNGKRSITRLDRAEFKFRQLLTPTHTATYMD
 VAELEIAIKVALVADRGAQVMEFADGVSRLSAGADDVRAEEDLVVDYAGEPLTIA
 FNPTLYTDGLSLSRSESVSFGFTTAGKPALEPRVSGDDRPVAGLNGNGFFPAVSTDY
 YLLMEVRUPG"
 3280. 4437
 /gene="recP"
 /locus_tag="Mb00003"
 3280. 4437
 /gene="recP"
 /locus_tag="Mb00003"
 /note="Mb00003, recP, len: 385 aa. Equivalent to Rv0003,
 len: 385 aa, from Mycobacterium tuberculosis strain H37Rv,
 (99.5% identity in 385 aa overlap). recP, DNA replication
 and repair protein (see citations below), equivalent to
 others Mycobacterial DNA replication and repair proteins
 e.g. NP_301131.1|NC_002677 from Mycobacterium leprae (385
 aa); Q9L7L5|RECF MYCFA from Mycobacterium avium subsp.
 paratuberculosis (385 aa); P50916|RECF MYCSM from
 Mycobacterium smegmatis (384 aa); etc. Also highly similar
 to others e.g. P36176|RECF_STRCO DNA REPLICATION AND
 REPAIR PROTEIN from Streptomyces coelicolor (373 aa);
 NP_440892.1|NC_000911 from Synecococcus sp. strain PCC
 6903 (384 aa); NP_469352.1|NC_003212 from Listeria innocua

(370 aa); etc. Contains P500017 ATP/GTP-binding site motif A (P-loop), P500617 RecF protein signature 1, and P500618 RecF protein signature 2. BELONGS TO THE REC F FAMILY."

/codon_start=1
/evidence=experimental
/transl_table=11
/product="DNA REPLICATION AND REPAIR PROTEIN REC F
(SINGLE-STRAND DNA BINDING PROTEIN)"
/protein_id="CAD92865.1"
/db_xref="GI:31616765"
/translation="MVVHGLGDFSRMACVLELHGRTFVFGNGYKTNLLEALW
YSTLGHRSYADLPILRVGTDRVITIVNDGRCVADLELATGRVKNRLRSSV
RSTRDVGVKURAFEDLGLVGPADRRYLDLAIIVRRPAIAAFAEYVVRVQ
RTALLKSVPGARVGRGSDTLEVDMSRLAEHVAARDLVNQLAPKVKAYQL
LAPESASIGYRSMQDVTGPSQSDTDRQLAARLLAALARRDELRGVCVLGPH
RDLILRLGQPAKGFASGEAWSLAVALRLAAYQLLRVDGGEPEVLLDDVFAELDM
RRALATAESAEQVLVTAALVEDIPAGWDARRVHIDVRAADTGTGSMVLP"
4434..4997
/locus_tag="Mb00004"
/locus_tag="Mb00004"
/note="Mb00004", len: 187 aa. Equivalent to Rv0004, len:
187 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.5% identity in 187 aa overlap). Conserved hypothetical
protein (see citation below). highly similar, but longer
21 aa in N-terminus, to AAF33696.1|AF222789 unknown
protein from Mycobacterium avium subsp. paratuberculosis
(166 aa); and highly similar to NP_301132.1|NC_002677
conserved hypothetical protein from Mycobacterium leprae
(189 aa); S70990 hypothetical protein from Mycobacterium
smegmatis (194 aa). Also similar to in C-terminus to
C-terminal part of P35925|YREG STRCO HYPOTHETICAL 19.8 KDA
PROTEIN (IN REC F-GRB INTERGENIC REGION) from Streptococcus
coelicolor (190 aa), FASTA scores: opt: 404, E(): 3.9e-18,
(40.7% identity in 189 aa overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD92866.1"
/db_xref="GI:31616766"
/translation="MTGSDVDRPDQNGERIMKSPGLDLVRLTDEARAARAGQDAG
RGVAVASGRVAGTSGSPGDPIDPQPLGKAARELAIRKGRWSVRVAGMVLGQW
SAVVGQIAHPTALPDNGVLVIAEASTAMATQLRIMQAQLAKIAAAGNDVVYRSL
KITGPAAPGWRKGPRIAGKGRPDYTG"
5123..7267
/gene="gyrB"
/locus_tag="Mb00005"
/locus_tag="Mb00005"
/EC_number="5.99.1.3"
/note="Mb00005, gyrB, len: 714 aa. Equivalent to Rv0005,
len: 714 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.9% identity in 714 aa overlap). gyrB, DNA gyrase
subunit B (EC 5.99.1.3) (see citations below), equivalent,
except in N-terminus, to other Mycobacterium DNA GYRASES
SUBUNIT B e.g. T10005 from Mycobacterium leprae (697 aa);
Q9L7L3|GYRB.MYCPA from Mycobacterium avium subsp.
paratuberculosis (677 aa) (has its N-terminus shorter);
P48355|GYRB.MYCSM from Mycobacterium smegmatis (675 aa);
etc. Also highly similar to others e.g. T10969 from
Streptococcus coelicolor (686 aa); P50075|GYBS_STRFH from

gene

CDS

Query Match 100.0%; Score 27; DB 1; Length 343050;

Best Local Similarity 100.0%; Pred. No. 0.063;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGAGACTTTCCTCCCAACCGAC 27

Db 55467 GTCGACGAGACTTTCCTCCCAACCGAC 55493

RESULT 5

NMU58765
LOCUS
DEFINITION

NMU58765 6972 bp DNA linear BCT 27-SEP-2002
Neisseria meningitidis trRNA gene, complete sequence; and beta 1,4
glucosyltransferase (lgTF), alpha 1,2 N-acetylglucosamine
transferase (lfaK), and putative sodium/chloride ion channel genes,
complete cds.

ACCESSION U58765 U35713
VERSION
KEYWORDS

SOURCE GI:4433816

ORGANISM

Neisseria meningitidis

Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;

Neisseriaceae; Neisseria.

REFERENCE 1 (bases 3576 to 4640)

AUTHORS

Kahler, C.M., Carlson, R.W., Rahman, M.M., Martin, L.E. and

Stephens, D.S.

TITLE Inner core biosynthesis of lipooligosaccharide (LOS) in Neisseria
meningitidis serogroup B: identification and role in LOS assembly
of the alpha1,2 N-acetylglucosamine transferase (RfaK)

J. Bacteriol. 178 (5), 1265-1273 (1996)

J. Bacteriol. 178 (5), 1265-1273 (1996)

REFERENCE 8631701

2 (bases 1 to 6972)

Kahler, C.M., Carlson, R.W., Rahman, M.M., Martin, L.E. and

Stephens, D.S.

TITLE Two glycosyltransferase genes, lgTF and rfaK, constitute the

lipooligosaccharide ice (inner core extension) biosynthesis operon

of Neisseria meningitidis

J. Bacteriol. 178 (23), 6677-6684 (1996)

J. Bacteriol. 178 (23), 6677-6684 (1996)

REFERENCE 97113422

3 (bases 1 to 6972)

Kahler, C.M., Carlson, R.W., Rahman, M.M., Martin, L.E. and

Stephens, D.S.

Direct Submission

Submitted (16-MAY-1996) Veterans Affairs Medical Center, 1670

Clairmont Rd, Research 151, Rm 5A161, Decatur, GA 30033, USA

4 (bases 1 to 6972)

Kahler, C.M., Carlson, R.W., Rahman, M.M., Martin, L.E. and

Stephens, D.S.

Direct Submission

Submitted (18-MAR-1999) Veterans Affairs Medical Center, 1670

Clairmont Rd, Research 151, Rm 5A161, Decatur, GA 30033, USA

REMARK Sequence update by submitter

COMMENT On Mar 18, 1999 this sequence version replaced gi:1698542.

FEATURES

source

1..6972

/organism="Neisseria meningitidis"

/mol_type="genomic DNA"

/strain="CDC8201085; NMB"

/db_xref="taxon:487"

repeat_region 53..148

/note="repeat element"

misc_feature 936..1532

/note="putative transposase"

repeat_region 936..1532

/insertion_seq="IS1016"

misc_feature <1534..2613

/note="putative beta-ketoacyl-ACP synthetase"

trna 2867..2742

gene 2817..3575

/gene="lgTF"

2817..3575

/gene="lgTF"

/function="lipooligosaccharide biosynthesis"

/codon_start=1

/transl_table=11

/product="beta 1,4 glucosyltransferase"

/protein_id="AAC44647.1"

/db_xref="GI:1698543"

/translation="MKKVSLLIVAKNEANHREICSRFDEKVIIVDDHSDNTAEI
AEGLGAKVFORHLNGDFGAKTFAEQAGGAVFLIDADERCTPELSISIKVQGD
YAAVFFVRRNLFENHPATHGAMRPDSVCRLMPKDSSEVGKHETVQTPYKRLKHP

MYHYVWNEQVFNKFNKTSISABKYEBOGKPVFVRDIIILRPINGFFKIYILNKGF
 LDGKMGWMSVNHSTYTIWKYKLYLKSGGKF
 3576. .4640
 /gene="rfak"
 3576. .4640
 /gene="rfak"
 /function="lipooligosaccharide biosynthesis"
 /codon_start=1
 /transl_table=11
 /product="alpha 1,2 N-acetylglucosamine transferase"
 /protein_id="AAC44648.1"
 /db_xref="GI:1209581"
 /translation="MEKEPRILNIVSAKINGGGEQVYDVSKALGLRGCTMTAVNKN
 NLMHRRPSEVSVTTTRHLTLNGLFSLCALTRFRNRIHLMHTGKIALSILK
 KUTGLVLIPEKNVANKYDFVHRLIQTGADVFVRSRLVVDVADNPFKEKRII
 NGIDTRPPSKDPSGPRFTVAYAGRISPEKLENLEACVILHRKYPQRLKLAGH
 GHPTDYMCRIDVSGASGAPFVSFEFTENIASFYRQSDVVLPSLVEAFGLSC
 MYCRTAVIGNTLGAQKEIVHEHQSGILLDRLTPESLADEIERLVNPEAKNALATAAH
 OCVAARFTINHTADKLDAI"
 4845. .5001
 /notes="repeat element"
 5028. .6365
 /notes="orfA"
 /codon_start=1
 /transl_table=11
 /product="putative sodium/chloride ion channel"
 /protein_id="AAD20893.1"
 /db_xref="GI:4433817"
 /translation="MSNHTSWSSKIGFVLAAGSAIGLGAIWKFPTAGTNGGAVFFL
 LFLFTLVALPVQLAEFYIGRTGKNVSPFVLRPGTQWLWGRMGVACFILLSE
 YSVGGWLVNYYHPSFTGAVHTGADFAFGALFATISNPAGSLSYQALFMLITVWYKGG
 ISDGIKANRYLMPGLFILFALAIKSLTLPAGMEGVSLPKPNWSYFKADMTITAGL
 QAPFALSIGVSAMITVASYLGKDDMRSGHTIMMNLVSLLAGLVIPPAVFARFGE
 PQCGGLIITVLPVAVMKMPFGTVLFAVFMILLVATLTSASMLETVIASIRQDER
 KRKHFWLIGTAIFVIGISALSGFVGKFKTIIDLDYIVISAVIMPGVAGVVA
 VFTAWIQDQKSVLDKAGAGSTVPRAVLLMLNKRKYLAPIAIIIVFVNSLDIL"
 6379. .6421
 /notes="hair-pin structure"
 ORIGIN
 Query Match 76.3%; Score 20.6; DB 1; Length 6972;
 Best Local Similarity 85.2%; Pred. No. 61;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GTCACAGCAGACTTCCCAACCCGAC 27
 |||||
 DB 6666 GTCAACGAGACTTCCGAGCCGAC 6692
 RESULT 6
 AE002521
 LOCUS
 DEFINITION
 Neisseria meningitidis serogroup B strain MC58 section 163 of 206
 of the complete genome.
 ACCESSION
 AE002521 AE002098
 VERSION
 AE002521.1 GI:7226962
 KEYWORDS
 SOURCE
 Neisseria meningitidis MC58
 ORGANISM
 Neisseria meningitidis MC58
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 1 (bases 1 to 14056)
 Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
 Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
 Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
 Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
 Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
 Cittiene,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
 Qian,H., Vamathevan,J., Gill,J., Scarlato,V., Masingani,V.,
 Pizzza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
 Rappuoli,R. and Venter,J.C.
 Complete genome sequence of Neisseria meningitidis serogroup B
 strain MC58

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 Science 287 (5459), 1809-1815 (2000)
 20175755
 10710307
 2 (bases 1 to 14056)
 Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
 Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
 Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
 Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
 Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
 Cittiene,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
 Qian,H., Vamathevan,J., Gill,J., Scarlato,V., Masingani,V.,
 Pizzza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
 Rappuoli,R. and Venter,J.C.
 Direct Submission
 Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1. 14056
 /organism="Neisseria meningitidis MC58"
 /mol_type="genomic DNA"
 /strain="MC58"
 /db_xref="taxon:122586"
 /notes="serogroup: B"
 120. .1457
 /gene="NMB1707"
 120. .1457
 /gene="NMB1707"
 /note="similar to GB:AL009126 percent identity: 75.06;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="sodium- and chloride-dependent transporter"
 /protein_id="AAF42055.1"
 /db_xref="GI:7226963"
 /translation="MSNHTSWSSKIGFVLAAGSAIGLGAIWKFPTAGTNGGAVFFL
 LFLFTLVALPVQLAEFYIGRTGKNVSPFVLRPGTQWLWGRMGVACFILLSE
 YSVGGWLVNYYHPSFTGAVHTGADFAFGALFATISNPAGSLSYQALFMLITVWYKGG
 ISDGIKANRYLMPGLFILFALAIKSLTLPAGMEGVSLPKPNWSYFKADMTITAGL
 QAPFALSIGVSAMITVASYLGKDDMRSGHTIMMNLVSLLAGLVIPPAVFARFGE
 PQCGGLIITVLPVAVMKMPFGTVLFAVFMILLVATLTSASMLETVIASIRQDER
 KRKHFWLIGTAIFVIGISALSGFVGKFKTIIDLDYIVISAVIMPGVAGVVA
 VFTAWIQDQKSVLDKAGAGSTVPRAVLLMLNKRKYLAPIAIIIVFVNSLDIL"
 6379. .6421
 /notes="hair-pin structure"
 ORIGIN
 Query Match 76.3%; Score 20.6; DB 1; Length 6972;
 Best Local Similarity 85.2%; Pred. No. 61;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GTCACAGCAGACTTCCCAACCCGAC 27
 |||||
 DB 6666 GTCAACGAGACTTCCGAGCCGAC 6692
 RESULT 6
 AE002521
 LOCUS
 DEFINITION
 Neisseria meningitidis serogroup B strain MC58 section 163 of 206
 of the complete genome.
 ACCESSION
 AE002521 AE002098
 VERSION
 AE002521.1 GI:7226962
 KEYWORDS
 SOURCE
 Neisseria meningitidis MC58
 ORGANISM
 Neisseria meningitidis MC58
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 1 (bases 1 to 14056)
 Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
 Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
 Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
 Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
 Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
 Cittiene,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
 Qian,H., Vamathevan,J., Gill,J., Scarlato,V., Masingani,V.,
 Pizzza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
 Rappuoli,R. and Venter,J.C.
 Complete genome sequence of Neisseria meningitidis serogroup B
 strain MC58

Science 287 (5459), 1809-1815 (2000)
 20175755
 10710307
 2 (bases 1 to 14056)
 Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
 Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
 Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
 Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
 Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
 Cittiene,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
 Qian,H., Vamathevan,J., Gill,J., Scarlato,V., Masingani,V.,
 Pizzza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
 Rappuoli,R. and Venter,J.C.
 Direct Submission
 Submitted (17-MAR-2000) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1. 14056
 /organism="Neisseria meningitidis MC58"
 /mol_type="genomic DNA"
 /strain="MC58"
 /db_xref="taxon:122586"
 /notes="serogroup: B"
 120. .1457
 /gene="NMB1707"
 120. .1457
 /gene="NMB1707"
 /note="similar to GB:AL009126 percent identity: 75.06;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="sodium- and chloride-dependent transporter"
 /protein_id="AAF42055.1"
 /db_xref="GI:7226963"
 /translation="MSNHTSWSSKIGFVLAAGSAIGLGAIWKFPTAGTNGGAVFFL
 LFLFTLVALPVQLAEFYIGRTGKNVSPFVLRPGTQWLWGRMGVACFILLSE
 YSVGGWLVNYYHPSFTGAVHTGADFAFGALFATISNPAGSLSYQALFMLITVWYKGG
 ISDGIKANRYLMPGLFILFALAIKSLTLPAGMEGVSLPKPNWSYFKADMTITAGL
 QAPFALSIGVSAMITVASYLGKDDMRSGHTIMMNLVSLLAGLVIPPAVFARFGE
 PQCGGLIITVLPVAVMKMPFGTVLFAVFMILLVATLTSASMLETVIASIRQDER
 KRKHFWLIGTAIFVIGISALSGFVGKFKTIIDLDYIVISAVIMPGVAGVVA
 VFTAWIQDQKSVLDKAGAGSTVPRAVLLMLNKRKYLAPIAIIIVFVNSLDIL"
 6379. .6421
 /notes="hair-pin structure"
 ORIGIN
 Query Match 76.3%; Score 20.6; DB 1; Length 6972;
 Best Local Similarity 85.2%; Pred. No. 61;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GTCACAGCAGACTTCCCAACCCGAC 27
 |||||
 DB 6666 GTCAACGAGACTTCCGAGCCGAC 6692
 RESULT 6
 AE002521
 LOCUS
 DEFINITION
 Neisseria meningitidis serogroup B strain MC58 section 163 of 206
 of the complete genome.
 ACCESSION
 AE002521 AE002098
 VERSION
 AE002521.1 GI:7226962
 KEYWORDS
 SOURCE
 Neisseria meningitidis MC58
 ORGANISM
 Neisseria meningitidis MC58
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.
 1 (bases 1 to 14056)
 Tettelin,H., Saunders,N.J., Heidelberg,J., Jeffries,A.C.,
 Nelson,K.E., Eisen,J.A., Ketchum,K.A., Hood,D.W., Peden,J.F.,
 Dodson,R.J., Nelson,W.C., Gwinn,M.L., DeBoy,R., Peterson,J.D.,
 Hickey,E.K., Haft,D.H., Salzberg,S.L., White,O., Fleischmann,R.D.,
 Dougherty,B.A., Mason,T., Ciecko,A., Parksey,D.S., Blair,E.,
 Cittiene,H., Clark,E.B., Cotton,M.D., Utterback,T.R., Khouri,H.,
 Qian,H., Vamathevan,J., Gill,J., Scarlato,V., Masingani,V.,
 Pizzza,M., Grandi,G., Sun,L., Smith,H.O., Fraser,C.M., Moxon,E.R.,
 Rappuoli,R. and Venter,J.C.
 Complete genome sequence of Neisseria meningitidis serogroup B
 strain MC58

MLQPNDSFGKRVLRISGNGVAAQAAEKAQIQAQKVLTVSDNSGVLFPDSDGWTAAQ
LAALIEKVRERVATYAEQGLQYFQKQPGVAEIALPCATQNEIDEEAAKTL
ANGCYVAEAGNPGSTGAGVQFIKAGILYAPKASNAGVATSGLEMSQNAIRLSWT
REEDQRLFGIMQSIHESCLKYGKVGDTVNYVANGANIAGFKVADAMLAQGP"
complement (4944. .5723)
/gene="NMB1711"
complement (4944. .5723)
/gene="NMB1711"
/note="similar to PID:455922 GB:U00096 SP:P06957
PID:1786303 PID:434009 percent identity: 53.25; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, GntR family"
/protein_id="AAF42058.1"
/db_xref="GI:7226966"
/translations="MKLVRRPKISDOVLSLEBRIAGVVAEGGKIPPERVLAERFVG
SRPSVRLAVNLVAROVLEARQDGYVSVKPODFLQSGWOELLGKHNSWEDVDPS
CHLEGWAAALAEERRTDADIKRIGFLEKFEACESENLHQSSEADYSHQTIADAAH
NLFLSHSGGLKMLRQTRSSLIYLNQESDPRPKMAQHRVLYEALSRRPGEASEA
AKAHLNVAASSILKREYQSRNRHDTLQNDLKRVDQWAV"
6154. .6453
/gene="NMB1712"
6154. .6453
/gene="NMB1712"
/note="similar to PID:882504 GB:U00096 SP:Q46939
PID:1789347 percent identity: 71.01; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="L-lactate permease-related protein"
/protein_id="AAP42059.1"
/db_xref="GI:7226967"
/translations="MRCFNWFDHCRLEKFKRCSIMETWQNTALGGSLYATAA
AALLPIVFFAALTVLKQYQAGLYTLIALAAVAFGFMPTGMVAVSLPPQPH"
complement (6839. .7804)
/gene="NMB1713"
complement (6839. .7804)
/gene="NMB1713"
/note="similar to SP:P37247 PID:1197006 percent identity:
61.94; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="transposase, IS30 family"
/protein_id="AAP42060.1"
/db_xref="GI:7226968"
/translations="MSYTOLOTGCRVHIOYLSRHCTVTEIAKOLARHKSITSEIRRH
RTQOQYSAEKAQRQRTIKORQPKYKDSQIIGHIDLIRKLSPEQVCAYLCKKH
RITLHSTIRYLRQDSNGSTLWOHLRCSKPYRKRYGSGTWGRKVPNRVGNRPA
IVDKRSIGDEWATTVGKQKSAITLVERTYRIICKLDSKAEDTARAVALRK
AKDRVHTITMDNGKEFYQHTKITKALAEYFCRPYHSEKGLNENGLIRQYFPK
QTDPRNISDREIRRVDELNRPRTKLTGYETPSVLFNLQPLIH"
complement (7914. .9317)
/gene="NMB1714"
complement (7914. .9317)
/note="similar to GB:L23839 PID:404773 PID:833173 percent
identity: 64.15; identified by sequence similarity;
putative"
/transl_table=11
/product="multidrug efflux pump channel protein"
/protein_id="AAP42061.1"
/db_xref="GI:7226969"
/translations="MDTKLTITLSVAAAFALSACTMIPQEQPKVEVAETFKNDTAD
SGIRAVLDGWHYFADPRQKLDITALENTSLRTAVLNSIYRKQYMIERNLLPTL
AANANDROGSLSGGNVSSYKGLCAASVELDLFCRVSSSEALQGVFASTNRDA
AHLJLTAYAKAFNERYAREASIAQRVLTRETYKLSRLRYKAGVISVALRQOE
ALIESKADYAHAREKREGRNALATLNPQIPEDLPAGLPDQKQFVKEKUPAGUSSE
VLDRPDTRAAEHALKQANVAGAAAFPSIRITGVGTGSBELGFLKSGTGWS
FAPSTPLPTFGTNGTNDVAKLQVQIVAYESAQVAFQDVAALAAAREGLDKAY
DALSKOSRASKALRLALRYKHVSGVGLDLDLDAERSYVAEAGALSQAQLTRAENLAD
LYKALGGGLKRDQTQDK"

gene
CDS

gene
CDS

gene
CDS

gene
CDS

gene
CDS

complement (9372. .12575)
/gene="NMB1715"
complement (9372. .12575)
/gene="NMB1715"
/note="similar to PID:1408202 percent identity: 98.59;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="multiple transferable resistance system protein
MrpD"
/protein_id="AAF42062.1"
/db_xref="GI:7226970"
/translations="MAKFFIDRPIFANVISITFIAGIFGKSLPVSQYPSVAAPTIT
LRATYPGASQVMEQSVLSVIERNMGVGLDYMSADSSGSGSVLTFPTDENL
ARVEYQNKLSVLSLTPATVOQYGVTVSKARSNFLIMVLSDDYQSTEENMDYQRNI
VPELQRIEYGQVRLFGAQRAMRTWDPKQNLQNLGFADVGSALSANQVISAQSIG
SLPVRGQTVATVTAQGLTAEEFGNVILRANTDGSNVYLDKVARVGLGMDYSIS
TRLGVNTGMVNLSSGNAMATAKAKERMATLEKYPQGSNKTPTDTSKFVEIS
IEKVIHTLIEAMVLFVVMVLFQNIIRYTLPTIIVFISLLCGFAPISYMGMSINVLIT
MFAMVLTIGVVDVAIVVENVERIMAGEGLPPKPEKAKMGQISGAVIGITAVLISV
FVPLAMTSGAGTNYKQFALTMASSIAFSLATLTLPALCATMLKTIPIKHHBEKKG
FGWFKFNSTHGYEGRVAKVLRKTRFMVVVYIGLAVGVFLFRLPSTFLPTEDQ
GFVMVSVQLPAGATQERTNATLAQVTLAKSIDPEIENITVSGFSGSGGNMAGFA
ILKQWNERTPGSDAVALAGTLGMMGLDKGFIAGVPPPILELNGSGSLSLNLOD
RNNTGHTALLAKRNELIQKWRASGLFDPSTVRAGLEDSPOLKIDINRAAAGGISF
ADIRLALASLSSYSVDFPNQRLQRMVQADEADARKQPADILNLTVPNKSQVAVPL
STIATVSENGTEQSVRFNPGYFNKLSASPAVSTGQAAVQKMWDELGGSYSLEW
GGSRBEAKGSGQTLILYGLAAVAVFLVLAALYESNIPILAVILVPLIGLIGAAVGT
GRNLFEGLLGSGVPSFANDIYFQGVFTVGLSNAKLIIEFARDLQAOQKSAVEAL
EALRFRPRTIMTSGFAFILGVPLIYIAGGASSASORAITVTWGLMGLTLLSLVLP
LFVYVVRKFKETAHEHMAVQAAEAGITGSDSDSQH"
complement (12587. .13825)

gene

Query Match 76.3%; Score 20.6; DB 1; Length 14056;
Best Local Similarity 85.2%; Pred. No. 66;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 GTCGAACGAGACTTCCCAACCGAC 27
DB 1758 GTCGAACGAGACTTCCCAACCGAC 1784

RESULT 7

NMA622491

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NMA622491 326301 bp DNA linear BCT 02-SEP-2002
Neisseria meningitidis serogroup A strain Z2491 complete genome;
segment 6/7.
ALI62757 ALI57959
ALI62757.2 GI:7380371
Neisseria meningitidis Z2491
Neisseria meningitidis Z2491
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
1 (bases 1 to 326301)
Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
Holroyd,S., Jagsels,K., Leather,S., Moule,S., Mungall,K.,
Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
Skelton,J., Whitehead,S., Spratt,B.G. and Barrall,B.G.
Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis Z2491
Nature 404 (6777), 502-506 (2000)
20222556
10761919
2 (bases 1 to 326301)
Parkhill,J.
Direct Submission
Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk
Notes:

Details of N. meningitidis sequencing at the Sanger Centre are available on the World Wide Web.

(URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).

FEATURES

```

source
1..326301
/organism="Neisseria meningitidis 22491"
/mol_type="genomic DNA"
/strain="22491"
/db_xref="taxon:122587"
/notes="serogroup: A"
repeat_unit
196..311
/notes="REP 2; hmms hit to HM REP 2 (1 - 109), score: 87.43"
repeat_unit
434..453
/notes">= 90% match to ATTCNNNNNNNGGAAT"
/label="dRS3"
repeat_unit
498..560
/notes="Correia element; hmms hit to HM Correia (1 - 62), score: 78.71"
/label="Correia"
repeat_unit
562..604
/notes="Correia element; hmms hit to HM Correia (114 - 156), score: 51.92"
/label="Correia"
complement(667..676)
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/misc_feature
/label=DUS
complement(668..1813)
/genes="dapE"
complement(668..1813)
/genes="dapE"
/EC_number="3.5.1.18"
/notes="NMA1730, dapE, succinyl-diaminopimelate desuccinylase, len: 381aa; similar to many eg. SW:P24176 (dape ECOLI) dape, succinyl-diaminopimelate desuccinylase from Escherichia coli (375 aa) fasta scores; E(): 0, 56.9% identity in 371 aa overlap. Contains Pfam match to entry PF01546 Peptidase_M20."
/codon_start=1
/transl_table=11
/product="succinyl-diaminopimelate desuccinylase"
/protein_id="CAB84958.1"
/db_xref="GI:7380372"
/db_xref="GOA:Q9JTL0"
/db_xref="SPTREMBL:Q9JTL0"
/translation="MTETQSLELAKALISRPSTVTPDRDCOKLLVERLYKIGFAAEEL
HGDTYKNILRGTKYVPVCFAGHTDVETGPEKWDSPPEPTERDCGLYGRGADM
KTSICFVACERFVAEHPDHQGSIALITSDDEGALDGTTKVDVLRKAGELIDYC
IVGEPTAVDKLDMKNGRGLSGNLTKVKGQHTAYPHLAIPNPHVTAPALLETQ
EVDENGTFFPTFSFQISNINGTGATNIPGELNVKFNFRFSTESTEAGLRQVHAI
LDKHGVYDLQWSCSQPFLTQAGKLTVDARAIAETCGIEAELSTTGTSQGRFIKA
IAKELIEGPSNATHIQINENVRLDDIPKLSAVYEGILARLLAGNAV"
complement(836..1786)
/genes="dapE"
/notes="Pfam match to entry PF01546 Peptidase_M20, , score 252.50, E-value 5.7e-72"
complement(1863..1872)
/misc_feature
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(1927..2547)
/genes="NMA1731"
complement(1927..2547)
/genes="NMA1731"
/notes="NMA1731, conserved hypothetical protein, len: 206aa; similar to SW:P44013 (V552 HAEIN) hypothetical protein from Haemophilus influenzae (207 aa) fasta scores; E(): 0, 59.6% identity in 203 aa overlap."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAB84959.1"
/db_xref="GI:7380373"
/db_xref="SPTREMBL:Q9JQZ3"

```

```

/translation="MLTPKSCDLNIPFFQFSQLKYPESIQIKADYKENWQIWOQ
LIQVAAELGAPFAPIHIERMCGWQVRAHFFAYFYKYQKNSAILLSILLNRRUSV
SLDMHCYKADVSPIALPDYNRWLDNFTDYKASFDMMHGAESYDDYRTVAQQNESDR
KLONDEDFPCIGKHIERDILGRQDVAKWIAETVEDLLPLYEACHGK"
complement(2551..2558)
/misc_feature
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS
complement(2581..3090)
/genes="NMA1732"
complement(2581..3090)
/genes="NMA1732"
/notes="NMA1732, conserved hypothetical protein, len: 169aa; similar to SW:P43971 (Y243 HAEIN) hypothetical protein from Haemophilus influenzae (172 aa) fasta scores; E(): 0, 49.7% identity in 169 aa overlap."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAB84960.1"
/db_xref="GI:7380374"
/db_xref="SPTREMBL:Q9JTK9"
/translation="MNPFTQSVTFAPSIEMLYACHGKVRRCQOVAMLSYIAENGC
NQLVQITRQITRYFNVAAPLHDEDEENFFLLQYAPQAESVDLLQHVSLHGN
WDVAASFAKLEADNAVIPDAEAFKRFVAGYDVHLATEEPLFDMGKMFIPKEKLEIG
EIMAA3RRR"
complement(3185..3736)
/genes="NMA1733"
complement(3185..3736)
/genes="NMA1733"
/notes="NMA1733, H.8 outer membrane protein, len: 183aa; strongly similar to SW:P07212 (H8 NEIME) H.8 outer membrane protein from Neisseria meningitidis (183 aa) fasta scores; E(): 0, 96.7% identity in 183 aa overlap. Contains Pfam match to entry PF00127 copper-bind, Copper binding proteins, plastocyanin/azurin family and Prosite match to PS00196 Type-1 copper (blue) proteins signature."
/codon_start=1
/transl_table=11
/product="H.8 outer membrane protein"
/protein_id="CAB84961.1"
/db_xref="GI:7380375"
/db_xref="GOA:P57025"
/db_xref="SWISS-PROT:P57025"
/translation="MKAYLALISAIVGLAACSOEPAAPAAEATPAEAPASEAPAAE
AAPADARAPAGNCAATVESNDNMONTKDIOYSKACKBPTITLKHTGTQPKTSMGH
NIVIGKVEDMDGIEKDGVGADTYVKPDDARVVAH"KLIGGESESLTLPKALDG
EYKFACTFPGHAGLMKGVTLVD"
complement(3191..3568)
/genes="NMA1733"
/notes="Pfam match to entry PF00127 copper-bind, Copper binding proteins, plastocyanin/azurin family, score 218.60, E-value 9.1e-62"
complement(3212..3259)
/misc_feature
/notes="PS00196 Type-1 copper (blue) proteins signature."
complement(4017..4439)
/genes="NMA1734"
complement(4017..4439)
/genes="NMA1734"
/notes="NMA1734, unknown, len: 140aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB84962.1"
/db_xref="GI:7380376"
/db_xref="SPTREMBL:Q9JTK8"
/translation="MNLQIGRRDALLSGLLKQAGQWRLLDAAVKLLPANLHPHFQCT
ACIEDGRLLVLLAANNMAARLKIAPSVIPOLAGLDASIRSVSVRLVPKPEKPKPTNT
LHLKAALESFSGSAVKLEKHPLEALANLVRHGA"
4306..4315
/misc_feature
/notes="Core DNA uptake sequence: gcgcgtctgaa"
/label=DUS

```

```

misc_feature complement(4486..4495)
/notes="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
gene 4570..7320
/genes="seca"
CDS 4570..7320
/genes="seca"
/notes="NMA1735, seca, preprotein translocase SecA subunit,
len: 916aa; similar to many eg. SW:PI0408 (SECA_ECOLI)
preprotein translocase SecA subunit from Escherichia coli
(901 aa) fasta scores; E(): 0, 58.0% identity in 920 aa
overlap. Contains Pfam match to entry PF01043
SecA protein, SecA protein, amino terminal region and
prosite match to PS01312 Protein seca signatures."
/codon_start=1
/transl_table=11
/product="preprotein translocase SecA subunit"
/protein_id="CAB84963.1"
/db_xref="GI:7380377"
/db_xref="GOA:O9JTK7"
/db_xref="SPTREMBL:O9JTK7"
/translation="MTNIAKIFGSRNRLKQYRKSVARINALEQNALSDADLQ
AKTAEKQRIADQDGLILPEAFVCREASRRITGMRHFDVOLIGMVLHDKIAEM
RTGEGTLVATLAVNLAKGKHVVTYNDYLSADGIMEFLYNFLGLTVGVISD
MQPFDQRNVAADITYGTNTEFGDYLRDMVTVDQYKQRELNFAVDEVDLSILIDE

Query Match 76.3%; Score 20.6; DB 1; Length 326301;
Best Local Similarity 85.2%; Pred. No. 94; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 4;

Qy 1 GTCGAACGAGACTTCCCAACCGAC 27
|||||
Db 234612 GTCGAACGAGACTTCCCAACCGAC 234638

RESULT 8
AX044033 349980 bp DNA linear PAT 24-NOV-2000
LOCUS
DEFINITION Sequence 112 from Patent WO0066791.
ACCESSION AX044033
VERSION AX044033.1 GI:11342917
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
Neisseriaceae; Neisseria.
REFERENCE
1
AUTHORS Pizza, M., Hickey, E., Peterson, J., Tettelin, H., Venter, J. C.,
Masignani, V., Galatetti, C., Mora, M., Ratti, G., Scariselli, M.,
Scariato, V., Rappuoli, R., Frazer, C. M. and Grandi, G.
TITLE Neisseria genomic sequences and methods of their use
JOURNAL
PATENT: WO 0066791-A 112 09-NOV-2000;
CHIRON CORPORATION (US); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1. 349980
/organism="Neisseria meningitidis"
/mol_type="unassigned DNA"
/db_xref="taxon:487"
/notes="sequence too long, cut in 8 pieces.-seq 1: 1 to
349980 349980 bases-seq 108: 300001 to 649980 349980
bases-seq 109: 600001 to 949980 349980 bases-seq 110:
900001 to 1249980 349980 bases-seq 111: 1200001 to 1549980
349980 bases-seq 112: 1500001 to 1849980 349980 bases-seq
113: 1800001 to 2149980 349980 bases-seq 114: 2100001 to
2272325 172325 bases"

ORIGIN
Query Match 76.3%; Score 20.6; DB 6; Length 349980;
Best Local Similarity 85.2%; Pred. No. 95; Indels 0;
Matches 23; Conservative 0; Mismatches 4; Gaps 0;

Qy 1 GTCGAACGAGACTTCCCAACCGAC 27
|||||

```

```

Db 284315 GTCGAACGAGACTTCCCAACCGAC 284341

RESULT 9
AX251053 6862 bp DNA linear PAT 05-OCT-2001
LOCUS
DEFINITION Sequence 21 from Patent WO0168912.
ACCESSION AX251053
VERSION AX251053.1 GI:15984476
KEYWORDS
SOURCE
ORGANISM
synthetic construct
artificial construct
artificial sequences.
REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with tumor suppressor genes and
oncogenes
JOURNAL
PATENT: WO 0168912-A 21 20-SEP-2001;
Epigenomics AG (DE)
FEATURES
source
1. 6862
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match 74.8%; Score 20.2; DB 6; Length 6862;
Best Local Similarity 88.0%; Pred. No. 97; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3;

Qy 2 TCGAACGAGACTTCCCAACCGAC 26
|||||
Db 6351 TCGAACGAGACTTCCCAACCGAC 6327

RESULT 10
AX251776 6862 bp DNA linear PAT 05-OCT-2001
LOCUS
DEFINITION Sequence 37 from Patent WO0168911.
ACCESSION AX251776
VERSION AX251776.1 GI:15985131
KEYWORDS
SOURCE
ORGANISM
synthetic construct
artificial construct
artificial sequences.
REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the cell cycle
JOURNAL
PATENT: WO 0168911-A 37 20-SEP-2001;
Epigenomics AG (DE)
FEATURES
Location/Qualifiers
source
1. 6862
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match 74.8%; Score 20.2; DB 6; Length 6862;
Best Local Similarity 88.0%; Pred. No. 97; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 3;

Qy 2 TCGAACGAGACTTCCCAACCGAC 26
|||||
Db 6351 TCGAACGAGACTTCCCAACCGAC 6327

RESULT 11
AX345124 6862 bp DNA linear PAT 01-FEB-2002
LOCUS
DEFINITION Sequence 195 from Patent WO0200928.
ACCESSION AX345124

```


VERSION AX345124.1 GI:18493010
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM artificial sequences.

REFERENCE 1
AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 020928-A 195 03-JAN-2002;
Epigenomics AG (DE)

FEATURES
source
1..5862
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match 74.8%; Score 20.2; DB 6; Length 6862;
Best Local Similarity 88.0%; Pred.No. 97;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCACACGAGACTTCCCAACCGA 26
||||||| ||||| ||||| ||||| |||||
Db 6351 TCACACGAGCTTTACCAACCGA 6327

RESULT 12
AC140072/c
LOCUS
DEFINITION Mus musculus chromosome UNK clone RP23-333A1, WORKING DRAFT
ACCESSION AC140072
VERSION AC140072.2 GI:28557991
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus (house mouse)

ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE The sequence of Mus musculus clone
JOURNAL Unpublished
REFERENCE 2
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
Submitted (20-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 3
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
Submitted (25-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
COMMENT On Feb 25, 2003 this sequence version replaced gi:28436310.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WGCSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
----- Project Information -----
Center project name: M_BA0333A01
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 179296 bases at least Q40
Consensus quality: 179715 bases at least Q30
Consensus quality: 179928 bases at least Q20
Insert size: 186000; agarose-fp

Insert size: 180534; sum-of-contigs
Quality coverage: 11.32 in Q20 bases; agarose-fp
Quality coverage: 11.66 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1163: contig of 1163 bp in length
* 1164 1263: gap of unknown length
* 1264 2698: contig of 1435 bp in length
* 2699 2798: gap of unknown length
* 2799 4207: contig of 1409 bp in length
* 4308 4307: gap of unknown length
* 4308 7295: contig of 2988 bp in length
* 7296 7395: gap of unknown length
* 7396 19384: contig of 11989 bp in length
* 19385 19485: gap of unknown length
* 19485 42541: contig of 23056 bp in length
* 42541 42641: gap of unknown length
* 42641 60918: contig of 18277 bp in length
* 60918 61018: gap of unknown length
* 61018 92352: contig of 31334 bp in length
* 92352 179886: contig of 87435 bp in length.
* 92452 Location/Qualifiers

FEATURES
source
1..179886
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-333A1"
1..1163
misc_feature
1264..2698
/note="assembly_name:Contig12"
misc_feature
2799..4207
/note="assembly_name:Contig22"
misc_feature
4308..7295
/note="assembly_name:Contig23"
misc_feature
7396..19384
/note="assembly_name:Contig24"
misc_feature
19485..42540
/note="assembly_name:Contig25"
misc_feature
42641..60917
/note="assembly_name:Contig26"
misc_feature
61018..92351
/note="assembly_name:Contig27"
misc_feature
92452..179886
/note="assembly_name:Contig28"
misc_feature
92452..179886
/note="assembly_name:Contig29"

ORIGIN
Query Match 71.1%; Score 19.2; DB 2; Length 179886;
Best Local Similarity 87.5%; Pred.No. 4.4e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCACACGAGACTTCCCAACCG 25
||||||| ||||| ||||| ||||| |||||
Db 94523 TCACACGAGACTTCCCAACCG 94500

RESULT 13
AL844840
LOCUS
DEFINITION Mouse DNA sequence from clone RP23-337P9 on chromosome 2, complete
AL844840
ACCESSION AL844840
VERSION AL844840.8 GI:31873535
KEYWORDS HTG.

SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 REFERENCE 1 (bases 1 to 219568)
 AUTHORS Tromans A.
 TITLE Direct Submission
 JOURNAL Submitted (17-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquary@sanger.ac.uk
 COMMENT On Jun 17, 2003 this sequence version replaced gi:31442058. Sequence from the Mouse Genome Sequencing Consortium whole genome shotgun may have been used to confirm this sequence. Sequence data from the whole genome shotgun alone has only been used where it has a phred quality of at least 30.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquary@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-337P9 is from the RPCI-23 Mouse BAC library

constructed by the group of Pieter de Jong.

For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6.

FEATURES
 source
 1..219568
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="2"
 /clone="RP23-337P9"
 /clone_lib="RPCI-23"

ORIGIN
 Query Match 71.1%; Score 19.2; DB 10; Length 219568;
 Best Local Similarity 87.5%; Pred. No. 4.5e+02;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGAGACTTCCCAACCG 25
 |||||
 DB 19415 TCGAACATGACTTCCCAACTG 19438

RESULT 14
 TNEGLA
 LOCUS 1382 bp DNA linear BCT 30-NOV-1997
 DEFINITION Thermotoga neapolitana bgla gene.
 ACCESSION 297212
 VERSION 297212.1 GI:2239255
 KEYWORDS beta-glucosidase; bgla gene.
 SOURCE Thermotoga neapolitana
 ORGANISM Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.

REFERENCE 1
 AUTHORS Dakhova, O.N., Kurepina, N.E., Zverlov, V.V., Svetlichnyi, V.A. and Velikodvorskaya, G.A.
 TITLE Cloning and expression in *Escherichia coli* of Thermotoga neapolitana genes coding for enzymes of carbohydrate substrate degradation
 JOURNAL Biochem. Biophys. Res. Commun. 194 (3), 1359-1364 (1993)
 MEDLINE 93356813
 PUBMED 8352795
 REMARK (sites)
 REFERENCE 2 (bases 1 to 1382)
 AUTHORS Zverlov, V.
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 1382)
 AUTHORS Zverlov, V.V.
 TITLE Direct Submission
 JOURNAL Submitted (30-JUN-1997) Zverlov V.V., Institute of Molecular Genetics, Russian Academy of Sciences, Kurchatov Sq. 46, Moscow, 123182, Russia
 FEATURES
 Location/Qualifiers
 1..1382
 /organism="Thermotoga neapolitana"
 /mol_type="genomic DNA"
 /strain="22706-MC24"
 /db_xref="taxon:2337"
 /clone="ptt6"
 96..100
 108..1382
 /gene="bgla"
 108..1382
 /gene="bgla"
 /EC number="3.2.1.21"
 /codon_start=1
 /evidence=experimental
 /translation="MKPEFGFLGVATASVOIEGSLADGAGWSIWHFTSPGNVK
 NGDTCGVADCHYRWKEDIIEIKIGAKYRFSISWPRILPECTGKVNQKGLDFYNR
 IDTLLEKNITTFITTYHWDLPFLSLKGGWANDIADWFAEYSKVLFEFGDRKRWI
 TLNPEWVAIVGHLYGVHAFPMKDIYVAFTVHNLRAHAKSVKVFRETVDGKIGIV
 FNGYFEPASEREEDIPARFMQFNYPFLNPIYRGEYPLVLEFAEYELPRNYD
 DMEIKOEIDFVGLNYYSGHWKYDPSPARVSNRNLPTAMGWEIVPEGYWILK
 GVKEEYNPOEYITENGAFFDDVYSEGKVDQNRIDYLRHTEQVWRAIQDGVPLKG
 YFVWSLLDNFEWAEGYSKRFGIYVVDYNTQKRI"

ORIGIN
 Query Match 70.4%; Score 19; DB 1; Length 1382;
 Best Local Similarity 81.5%; Pred. No. 3.2e+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GTCGACGAGACTTCCCAACCGAC 27
 |||||
 DB 1038 CTCGAGAGAACCTTCCCAACCGCC 1064

RESULT 15
 AKI10440/c
 LOCUS 3512 bp mRNA linear PLN 24-JUL-2003
 DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:002-166-D03, full insert sequence.
 ACCESSION AKI10440
 VERSION AKI10440.1 GI:32995649
 KEYWORDS FLI CDNA; oligo capping.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1

AUTHORS

The Rice Full-Length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team:
 Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yasaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group: Otsuno, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narioka, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Niikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN: Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Arakawa, T., Fukuda, S., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Yoshino, M., and Hayashizaki, Y.

TITLE Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)

MEDLINE 2752273

PUBMED 12869764

REFERENCE 2 (bases 1 to 3512)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narioka, R., Niikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Otsuno, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shingawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yasaki, J., Yokomizo, S. and Yoshimura, A.

TITLE

Direct Submission
 Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression: 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp,
 Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT This clone is one of the 28K full-length cDNA clones from japonica rice.

URL : <http://cdna01.dna.affrc.go.jp/cdna/>
 NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yasaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Otsuno, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narioka, R., Niikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.
 Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,

Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

FEATURES

source

1. 3512
 Location/Qualifiers
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="mRNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /clone="002-166-D03"

ORIGIN

Query Match 70.4%; Score 19; DB 8; Length 3512;

Best Local Similarity 81.5%; Pred. No. 3.5e+02;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GTCGAACGAGACTTTCCTCCCAACCGAC 27

|||||

Db 1083 GTCGATCGCGATTTCCTCCCAACCGAC 1057

|||||

Search completed: April 29, 2004, 06:01:22

Job time : 419.735 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:01:30 ; Search time 2128.98 Seconds
(without alignments)
378.717 Million cell updates/sec

Title: US-10-624-714-1
Perfect score: 27
Sequence: 1. gtcgaacgagactttccccaacacgac 27

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pin:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.8	73.3	230	29	CE197512 tigr-gss-
C 2	19.6	72.6	499	28	CC437866 PUHNS77TD
C 3	19.6	72.6	549	12	BI747890 rm47a09.y
C 4	19.6	72.6	842	28	BZ991196 PUGHQ22TB

5	19.6	72.6	1007	12	BG865804
6	19.4	71.9	646	12	BI235377
c 7	19.2	71.1	565	12	BI506842
8	19	70.4	472	12	BI896083
9	19	70.4	491	12	BG544187
10	19	70.4	651	28	BZ657986
11	19	70.4	686	29	CC620544
c 12	19	70.4	722	11	CNS0978E
c 13	19	70.4	798	29	CG377025
c 14	19	70.4	811	28	BZ830561
c 15	19	70.4	853	29	CG377037
c 16	19	70.4	898	10	BE584451
17	19	70.4	933	29	CG711110
18	18.8	69.6	455	12	BG098247
19	18.6	68.9	645	10	BF646171
20	18.6	68.9	803	14	CB238319
21	18.6	68.9	809	14	CA466776
c 22	18.6	68.9	865	14	CF551986
c 23	18.6	68.9	1049	9	AL572513
24	18.2	67.4	465	28	AZ249689
25	18.2	67.4	618	28	BH764137
c 26	18.2	67.4	681	28	BH188053
c 27	18.2	67.4	681	29	CNS0785U
c 28	18.2	67.4	791	28	BH508894
29	18.2	67.4	795	14	CD109689
30	18.2	67.4	814	29	CG736008
31	18.2	67.4	920	28	BZ793807
c 32	18.2	67.4	1426	29	AG034573
c 33	18	66.7	111	11	CNS088TQ
c 34	18	66.7	412	28	CC112049
c 35	18	66.7	478	12	BU161735
c 36	18	66.7	510	14	CF882673
c 37	18	66.7	520	28	BZ437910
c 38	18	66.7	561	12	BJ157080
c 39	18	66.7	591	14	CA184852
40	18	66.7	602	14	CB175627
41	18	66.7	608	29	CC966752
42	18	66.7	643	14	CB175329
c 43	18	66.7	644	12	BG966722
c 44	18	66.7	644	14	CB276044
c 45	18	66.7	667	12	BM612474

ALIGNMENTS

RESULT 1
CE197512/c
LOCUS tigr-gss-dog-17000372163904 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE197512
VERSION CE197512.1 GI:35353165
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
1 (bases 1 to 230)
Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
Venter,J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
22875432
PUBMED 14512627
COMMENT Contact: Kirknes EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org

```

Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..230
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /notes="Site 1: BstXI; Libraries were prepared from
                     peripheral blood"

ORIGIN
Query Match          73.3%; Score 19.8; DB 29; Length 230;
Best Local Similarity 91.3%; Pred. No. 3.4e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 GAACGAGACTTCCCAACCGA 26
    |||||
Db 56 GAACGAGACTTCCCAACCGA 34

RESULT 2
CC437866             499 bp      DNA      linear      GSS 20-MAY-2003
LOCUS                PUHNS777D ZM_0.6_1.0_KB Zea mays genomic clone ZMMETA524M10,
DEFINITION           genomic survey sequence.
ACCESSION            CC437866
VERSION              CC437866.1 GI:30936392
KEYWORDS             GSS.
SOURCE               Zea mays
ORGANISM             Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 499)
Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Reinick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
Maize Genomics Consortium
Unpublished (2003)
Other GSSs: PUHNS77B
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..499
                     /organism="Zea mays"
                     /mol_type="genomic DNA"
                     /strain="B73"
                     /db_xref="taxon:4577"
                     /clone="ZMMETA524M10"
                     /clone_lib="ZM_0.6_1.0_KB"
                     /notes="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb kb
                     Cot selected genomic DNA library"

ORIGIN
Query Match          72.6%; Score 19.6; DB 28; Length 499;
Best Local Similarity 84.6%; Pred. No. 5e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 TCGAACGAGACTTCCCAACCGAC 27
    |||||
Db 480 TCGAACGAGACTTCCCAACCGAC 455

RESULT 3
BI747890
LOCUS                BZ991196
DEFINITION           PUGHQ22TB ZM_0.6_1.0_KB Zea mays genomic clone ZMMETA386D19,
                     genomic survey sequence.
ACCESSION            BZ991196
VERSION              BZ991196.1 GI:29232751
KEYWORDS             GSS.
SOURCE               Zea mays

```

DEFINITION

rm47a09.y1 Meloioogyne arenaria egg pAMP1 v1 Chiapelli McCarter
Meloioogyne arenaria cDNA 5' similar to SW:ACHX ONCVO P54247
ACETYLCHOLINE RECEPTOR PROTEIN, NON-ALPHA CHAIN", mRNA sequence.

ACCESSION BI747890
VERSION BI747890.1 GI:15769692

KEYWORDS

EST.

SOURCE

Meloioogyne arenaria

ORGANISM

Meloioogyne arenaria

Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heteroderidae; Meloioogyminae; Meloioogyne.

REFERENCE

1 (bases 1 to 549)

AUTHORS

McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J.,
Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B.,
Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C.,
Tsagarashvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C.,
Underwood,K., Steptoe,M., Allen,M., Petson,B., Swaller,T.,
Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
McCann,R., Waterston,R. and Wilson,R.
The Washington Univ. Nematode EST Project, 1999

TITLE

Contact: McCarter JP

JOURNAL

Unpublished (1999)

COMMENT

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

The library was constructed by Brandi Chiapelli and Dr. James

McCarter (bchiapell@watson.wustl.edu & jmcarter@watson.wustl.edu) at

Washington University, St. Louis. DNA Sequencing by: Washington

University Genome Sequencing Center St. Louis. Nematodes were

provided by Dr. David Bird and Daniel Snyder of North Carolina

State University.

Seq primer: -40RP from Gibco

High quality sequence stop: 415.

Location/Qualifiers

1..549

/organism="Meloioogyne arenaria"

/mol_type="mRNA"

/db_xref="taxon:6304"

/dev_stage="egg"

/lab_host="DH10B"

/clone_lib="Meloioogyne arenaria egg pAMP1 v1 Chiapelli

McCarter"

/note="Vector: pAMP1 (Gibco); Site 1: NotI; Site 2: SalI;

The library was constructed by Brandi Chiapelli and Dr.

James McCarter at Washington University, St. Louis. The

cDNA was made by using Dynabead oligo-dT priming (Dynal).

PCR based library using a modified protocol from the

SMART PCR cDNA Synthesis Kit from Clontech. Directionally

cloned into the UDG sites of pAMP1. Nematodes were

provided by Dr. David Bird and Daniel Snyder of North

Carolina State University."

ORIGIN

Query Match 72.6%; Score 19.6; DB 12; Length 549;

Best Local Similarity 84.6%; Pred. No. 5.1e+02;

Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 TCGAACGAGACTTCCCAACCGAC 27

|||||

Db 160 TTGAACGTGACTTTCCTCCCAATCGGC 185

|||||

RESULT 4

BZ991196/c

LOCUS BZ991196

DEFINITION PUGHQ22TB ZM_0.6_1.0_KB Zea mays genomic clone ZMMETA386D19,

genomic survey sequence.

ACCESSION BZ991196

VERSION BZ991196.1 GI:29232751

KEYWORDS GSS.

SOURCE Zea mays

```

ORGANISM      Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE     1 (bases 1 to 842)
AUTHORS      Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
TITLE        Maize Genomics Consortium
JOURNAL      Unpublished (2003)
COMMENT      Other GSSs: FUGHQ22TD
Contact: Cathy Whitelaw
TIGR         9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES     Location/Qualifiers
source       1..842
             /organism="Zea mays"
             /mol_type="genomic DNA"
             /strain="B73"
             /db_xref="taxon:4577"
             /clone="ZMMBTA386D19"
             /clone_lib="ZM.0.5.1.0 KB"
             /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
Cor selected genomic DNA library"
ORIGIN
Query Match      72.6%; Score 19.6; DB 28; Length 842;
Best Local Similarity 84.6%; Pred. No. 5.7e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 TCGACGACGACTTTCCTCCCAACCGAC 27
      |||||
DB      403 TCGACACAGACTTGCCCAACCGGC 378

RESULT 5
BG865804      1007 bp mRNA linear EST 29-MAY-2001
LOCUS        602798202F1 NCI_CGAP_SG2 Mus musculus cDNA clone IMAGE:4914000 5',
DEFINITION   mRNA sequence.
ACCESSION    BG865804
VERSION      BG865804.1 GI:14216344
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE    1 (bases 1 to 1007)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgabps@mail.nih.gov
             Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10820 row: d column: 01
             High quality sequence stop: 712.

FEATURES     Location/Qualifiers
source       1..1007
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="FVB/N"
             /db_xref="taxon:10090"

```

```

/clone="IMAGE:4914000"
/lab_host="DHI0B (T1 phage-resistant)"
/clone_lib="NCI_CGAP_SG2"
/note="Organ: salivary gland; Vector: pCMV-SPORT6; Site 1:
NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo
dT. Average insert size 1.3 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
ORIGIN
Query Match      72.6%; Score 19.6; DB 12; Length 1007;
Best Local Similarity 84.6%; Pred. No. 6e+02;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 TCGAACGACGACTTTCCTCCCAACCGAC 27
      |||||
DB      154 TCGAATGACAGACGTCCCAACAGAC 179

RESULT 6
BI235377
LOCUS        RE31217.5prime RE Drosophila melanogaster normalized Embryo pFLC-1
DEFINITION   Drosophila melanogaster cDNA clone RE31217.5 similar to CG5765;
FBan0005765 located on: 2R 55B1-55B1; 05/12/2001, mRNA sequence.
BI235377
ACCESSION    BI235377.1 GI:14703820
VERSION      BI235377
KEYWORDS     EST.
SOURCE       Drosophila melanogaster (fruit fly)
ORGANISM     Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE    1 (bases 1 to 646)
AUTHORS      Stapleton,M., Brokstein,P., Hong,L., Tyler,D., Berman,B.,
Carlson,J., Champe,M., Chavez,C., Dorsett,V., Farfan,D., Frise,E.,
George,R., Gonzalez,M., Guarin,H., Harris,N., Li,P., Liao,G.,
Misra,S., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
Phouanavong,S., Wan,X., Yu,C., Lewis,S.E., Celniker,S. and
Rubin,G.M.
TITLE        BDGP/HHMI RE Drosophila EST Project
JOURNAL      Unpublished (2001)
COMMENT      Contact: Stapleton, M.
             BDGP
             Lawrence Berkeley National Lab
             One Cyclotron Rd, Berkeley, CA 94720, USA
             Fax: 510 486 6798
             Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
             hit genomic AE003801: arm:2R [12763208,13035857]
             estimated-cyto:54Fl-55B2: 05/12/2001
             Plate: RE.312 row: B column: 5
             High quality sequence stop: 559.
             Location/Qualifiers
             1..646
             /organism="Drosophila melanogaster"
             /mol_type="mRNA"
             /db_xref="taxon:7227"
             /clone="RE31217"
             /sex="male and female"
             /dev_stage="0-24 hours mixed stage embryonic"
             /lab_host="DH5-alpha Tona"
             /clone_lib="RE Drosophila melanogaster normalized Embryo
pFLC-1"
             /note="Organ: embryo; Vector: pFLC1; Site 1: XhoI; Site 2:
BamHI; Library was kindly generated by Piero Carninci at
the RIKEN. The library was normalized and excised using
Cre recombinase. Plasmid cDNA library."
ORIGIN
Query Match      71.9%; Score 19.4; DB 12; Length 646;
Best Local Similarity 95.2%; Pred. No. 6.5e+02;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      7 CGAGACTTTCCTCCCAACCGAC 27

```


CNS097AE	722 bp	linear	HTC 07-JAN-2003
Single read from an extremity of a full-length cDNA clone made from mRNA			

Anopheles gambiae total adult females. 5-PRIME end of clone
 FK0AAC19CH10 of strain 6-9 of Anopheles gambiae (African malaria
 mosquito).
 ACCESSION BX045522 GI:27618803
 VERSION HTG
 KEYWORDS Anopheles gambiae (African malaria mosquito)
 SOURCE Anopheles gambiae
 ORGANISM Anopheles gambiae
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
 Anopheles.
 1 (bases 1 to 722)
 Genoscope.
 Direct Submission
 TITLE Submitted (06-JAN-2003) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 FEATURES Location/Qualifiers
 source 1..722
 /organism="Anopheles gambiae"
 /mol_type="mRNA"
 /strain="6-9"
 /db_xref="taxon:7165"
 /clone="FK0AAC19CH10"
 /plasmid="pME18S-FL"
 /notes="end : 5-PRIME"

ORIGIN

Query Match 70.4%; Score 19; DB 11; Length 722;
 Best Local Similarity 81.5%; Pred. No. 1e+03; 5; Indels 0; Gaps 0;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GTCGACGAGACTTCCCAACCGAC 27

Db 614 GTCGACGAGAAATTTGCCCAACCGAC 588

RESULT 13

CG377025 798 bp DNA linear GSS 26-AUG-2003
 LOCUS OXGA87TV ZM_0.7_1.5_KB Zea mays genomic clone ZM85A0585005,
 genomic survey sequence.

ACCESSION CG377025

VERSION CG377025.1 GI:34294292

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 798)

REFERENCE

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,

Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,

Citek,R.W., Nurnberg,A., Robbins,D. and Lakey,N.

Consortium for Maize Genomics

Unpublished (2002)

Other GSSs: OXGA87TV

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

FEATURES

source 1..798
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZM85A0585005"
 /clone_lib="ZM_0.7_1.5_KB"

/note="Vector: pCSK-; Site 1: HincII; 0.7-1.5 kb
 methylation filtered genomic DNA library"

ORIGIN

Query Match 70.4%; Score 19; DB 29; Length 798;
 Best Local Similarity 81.5%; Pred. No. 1e+03; 5; Indels 0; Gaps 0;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GTCGACGAGACTTCCCAACCGAC 27

Db 370 GCGACCAAGACATGCCCAACCGAC 396

RESULT 14

BZ830561/c
 LOCUS PUFJ07TB ZM_0.6_1.0_KB Zea mays genomic clone ZM85A0358B14,
 genomic survey sequence.

ACCESSION BZ830561

VERSION BZ830561.1 GI:29054256

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 811)

REFERENCE

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,

Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and

Bennetzen,J.

Maize Genomics Consortium

Unpublished (2003)

Other GSSs: PUFJ07TD

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

FEATURES

source 1..811
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZM85A0358B14"
 /clone_lib="ZM_0.6_1.0_KB"
 /note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
 Cot selected genomic DNA library"

ORIGIN

Query Match 70.4%; Score 19; DB 28; Length 811;
 Best Local Similarity 81.5%; Pred. No. 1e+03;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GTCGACGAGACTTCCCAACCGAC 27

Db 524 GCGACCAAGACATGCCCAACCGAC 498

RESULT 15

CG377037/c
 LOCUS OXGA87TV ZM_0.7_1.5_KB Zea mays genomic clone ZM85A0585005,
 genomic survey sequence.

ACCESSION CG377037

VERSION CG377037.1 GI:34294304

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Rudiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)

TITLE JOURNAL COMMENT

Other_GSSs: OGA587TH
Contact: Cathy Whitelaw

TIGR Medical Center Drive, Rockville, MD 20850, USA
9712
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@cigr.org
Seq Primer: TF

FEATURES

source

Class: sheared ends.
Location/Qualifiers
1..853
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM585005"
/clone_lib="ZM 0.7 1.5 kb"
/note="Vector: pBCSK-; Site_1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN

Query Match 70.4%; Score 19; DB 29; Length 853;
Best Local Similarity 81.5%; Pred. No. 1e+03;
Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GTCGAACGAGACTTCCCAACCGAC 27
DB 674 GCGACCAAGACATGCCCAACCGAC 648

Search completed: April 29, 2004, 11:36:41
Job time : 2134.98 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:13:15 ; Search time 47.6471 Seconds
(without alignments)
314.472 Million cell updates/sec

Title: US-10-624-714-1

Perfect score: 27
Sequence: 1 gtcgaacgagactttccccaacccgac 27

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:
5: /cgn2_6/prodata/2/ina/6C_COMB.seq:
6: /cgn2_6/prodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appli
2	27	100.0	4411529	3 US-09-103-840A-1	Sequence 1, Appli
3	17.4	64.4	1230025	4 US-09-198-452A-1	Sequence 1, Appli
4	17.2	63.7	514	4 US-09-404-879A-251	Sequence 251, App
5	17.2	63.7	514	4 US-09-338-933-251	Sequence 251, App
6	17.2	63.7	514	4 US-09-215-681-251	Sequence 251, App
7	17.2	63.7	514	4 US-09-216-003A-251	Sequence 251, App
8	17	63.0	2394	4 US-09-994-192-3	Sequence 3, Appli
9	17	63.0	2394	4 US-10-181-660-2	Sequence 2, Appli
10	17	63.0	2394	4 US-10-181-660-10	Sequence 10, Appli
11	17	63.0	5333	4 US-08-956-171B-170	Sequence 170, App
12	16.4	60.7	933	2 US-08-701-191A-4	Sequence 4, Appli
13	16.4	60.7	933	4 US-09-664-526-4	Sequence 4, Appli
14	16.4	60.7	1056	2 US-08-701-191A-5	Sequence 5, Appli
15	16.4	60.7	1056	4 US-09-664-526-5	Sequence 5, Appli
16	16.4	60.7	1317	4 US-09-252-991A-11011	Sequence 11011, A
17	16.4	60.7	1782	4 US-09-252-991A-15714	Sequence 15714, A
18	16.4	60.7	2360	4 US-09-023-655-1288	Sequence 1288, Ap
19	16.4	60.7	2459	1 US-07-997-133-2	Sequence 2, Appli
20	16.4	60.7	2459	1 US-08-459-296-1	Sequence 1, Appli
21	16.4	60.7	2459	5 US-07-997-133-2	Sequence 2, Appli
22	16.4	60.7	2601	4 US-09-252-991A-11077	Sequence 11077, A
23	16.4	60.7	2610	4 US-09-252-991A-11177	Sequence 11177, A
24	16.4	60.7	2662	2 US-08-451-822A-14	Sequence 14, Appl
25	16.4	60.7	2662	4 US-08-323-430-14	Sequence 14, Appl
26	16.4	60.7	2733	1 US-08-371-001-14	Sequence 14, Appl
27	16.4	60.7	2733	5 PCT-US96-00331-14	Sequence 14, Appl

C	28	16.4	60.7	3132	4 US-09-252-991A-15639	Sequence 15639, A
	29	16.4	60.7	4887	1 US-07-789-915A-7	Sequence 7, Appli
	30	16.4	60.7	4887	1 US-08-005-002C-7	Sequence 7, Appli
	31	16.4	60.7	4887	1 US-08-487-203A-7	Sequence 7, Appli
C	32	16.4	60.7	536165	4 US-09-214-808-1	Sequence 1, Appli
C	33	16.4	60.7	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C	34	16.4	60.7	4411529	3 US-09-103-840A-1	Sequence 1, Appli
	35	16	59.3	1638	2 US-08-541-759B-1	Sequence 1, Appli
	36	16	59.3	5769	4 US-09-668-673B-1	Sequence 1, Appli
	37	15.8	58.5	234	4 US-09-252-991A-4843	Sequence 4843, Ap
	38	15.8	58.5	417	4 US-09-894-844-16	Sequence 16, Appl
	39	15.8	58.5	564	4 US-09-252-991A-13912	Sequence 13912, A
C	40	15.8	58.5	621	3 US-08-998-416-130	Sequence 130, App
	41	15.8	58.5	643	4 US-09-669-751-75	Sequence 75, Appl
C	42	15.8	58.5	932	1 US-08-458-912-1	Sequence 1, Appli
C	43	15.8	58.5	932	1 US-08-461-179-1	Sequence 1, Appli
C	44	15.8	58.5	932	1 US-08-459-254-1	Sequence 1, Appli
C	45	15.8	58.5	932	1 US-08-459-255-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 100.0%; Score 27; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred.No. 0.0046;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGACGAGACTTTCCTCCCAACCGAC 27
DB 55423 GTCGACGAGACTTTCCTCCCAACCGAC 55449

RESULT 2
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match      100.0%; Score 27; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. NO. 0.0048;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  GTGACGACGACTTCCCAACCGAC 27
Db      55481 GTGACGACGACTTCCCAACCGAC 55507

RESULT 3
US-09-198-452A-1/c
; Sequence 1, Application US/09198452A
; Patent No. 6552294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (30001)..(45000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (45001)..(60000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (60001)..(75000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (75001)..(90000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (90001)..(105000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (105001)..(120000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (120001)..(135000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (135001)..(150000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (150001)..(165000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (165001)..(180000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (180001)..(195000)
; OTHER INFORMATION: n=a or c or g or t

; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (195001)..(210000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (210001)..(225000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (225001)..(240000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (240001)..(255000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (255001)..(270000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (270001)..(285000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (285001)..(300000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (300001)..(315000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (315001)..(330000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (330001)..(345000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (345001)..(360000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (360001)..(375000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (375001)..(390000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (390001)..(405000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (405001)..(420000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (420001)..(435000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (435001)..(450000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (450001)..(465000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (465001)..(480000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (480001)..(495000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (495001)..(510000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (510001)..(525000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (525001)..(540000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (540001)..(555000)
; OTHER INFORMATION: n=a or c or g or t
```

```
; NAME/KEY: misc_feature
; LOCATION: (555001)..(570000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (570001)..(585000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (585001)..(600000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (600001)..(615000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (615001)..(630000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (630001)..(645000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (645001)..(660000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (660001)..(675000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (675001)..(690000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (690001)..(705000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (705001)..(720000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (720001)..(735000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (735001)..(750000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (750001)..(765000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (765001)..(780000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (780001)..(795000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (795001)..(810000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (810001)..(825000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (825001)..(840000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (840001)..(855000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (855001)..(870000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (870001)..(885000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (885001)..(900000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (900001)..(915000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
```

```
Query Match 54.4%; Score 17.4; DB 4; Length 1230025;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GTCGACGAGACTTCCCAACCGAC 27
      ||||| ||||| ||||| |||||
Db 270498 GTCGAAGAGCTATGCCAAGCGAC 270472
```

```
RESULT 4
US-09-404-879A-251
; Sequence 251, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 251
; LENGTH: 514
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(514)
; OTHER INFORMATION: n = A,T,C or G
US-09-404-879A-251
```

```
Query Match 63.7%; Score 17.2; DB 4; Length 514;
Best Local Similarity 86.4%; Pred. No. 38;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 2 TCGAACGAGACTTCCCAAC 23
      ||||| ||||| ||||| |||||
Db 482 TCACACAAACTTCCCAAC 503
```

```
RESULT 5
US-09-338-933-251
; Sequence 251, Application US/09338933
; Patent No. 6488931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
; FILE OF INVENTION: OVARIAN CANCER
; FILE REFERENCE: 210121.462C1
; CURRENT APPLICATION NUMBER: US/09/338,933
; CURRENT FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 251
; LENGTH: 514
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(514)
; OTHER INFORMATION: n = A,T,C or G
US-09-338-933-251
```

```
Query Match 63.7%; Score 17.2; DB 4; Length 514;
Best Local Similarity 86.4%; Pred. No. 38;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 2 TCGAACGAGACTTCCCAAC 23
```

Db 482 TCAACAAAGACTTCCCAAC 503

RESULT 6

US-09-215-681-251
; Sequence 251, Application US/09215681A
; Patent No. 6528253
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Frudakis, Tony N.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
; FILE REFERENCE: 210121.463
; CURRENT APPLICATION NUMBER: US/09/215,681A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 310
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 251
; LENGTH: 514
; TYPE: DNA
; ORGANISM: Homo sapien
; NAME/KEY: misc feature
; LOCATION: (1)...(514)
; OTHER INFORMATION: n = A,T,C or G
US-09-215-681-251

Query Match 63.7%; Score 17.2; DB 4; Length 514;
Best Local Similarity 86.4%; Pred. No. 38;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGAGACTTCCCAAC 23
Db 482 TCAACAAAGACTTCCCAAC 503

RESULT 7

US-09-216-003A-251
; Sequence 251, Application US/09216003A
; Patent No. 6670463
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Frudakis, Tony N.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF OVARIAN CANCER
; FILE REFERENCE: 210121.462
; CURRENT APPLICATION NUMBER: US/09/216,003A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 310
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 251
; LENGTH: 514
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: modified_base
; LOCATION: (8)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (107)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (312)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (338)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (351)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base

; LOCATION: (352)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (357)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (363)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (366)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (373)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (380)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (405)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (421)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (424)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (508)
; OTHER INFORMATION: Where n is a, c, g or t
US-09-216-003A-251

Query Match 63.7%; Score 17.2; DB 4; Length 514;
Best Local Similarity 86.4%; Pred. No. 38;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGAGACTTCCCAAC 23
Db 482 TCAACAAAGACTTCCCAAC 503

RESULT 8

US-09-994-192-3/c
; Sequence 3, Application US/09994192
; Patent No. 6610306
; GENERAL INFORMATION:
; APPLICANT: Judd, Ralph C.
; APPLICANT: Manning, Scott D.
; TITLE OF INVENTION: Omp85 Proteins of Neisseria gonorrhoeae and Neisseria meningitidis
; FILE REFERENCE: UMSB147AUSA
; CURRENT APPLICATION NUMBER: US/09/994,192
; CURRENT FILING DATE: 2001-11-26
; PRIOR APPLICATION NUMBER: US 09/177,039
; PRIOR FILING DATE: 1998-10-22
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2394
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2391)
; OTHER INFORMATION:
US-09-994-192-3

Query Match 63.0%; Score 17; DB 4; Length 2394;
Best Local Similarity 80.0%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TCGAACGAGACTTCCCAACCGA 26
Db 818 TCGATGAGACTTTCGCCCAACGA 794


```

RESULT 9
US-10-181-660-2/c
; Sequence 2, Application US/10181660
; Patent No. 6655956
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: SUPPLEMENTED OMV VACCINE AGAINST MENINGOCOCCUS
; FILE REFERENCE: P023785W0
; CURRENT APPLICATION NUMBER: US/10/181,660
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: GB-0001067.8
; PRIOR FILING DATE: 2000-01-17
; PRIOR APPLICATION NUMBER: GB-0005699.4
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 2
; LENGTH: 2394
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-10-181-660-2
Query Match 63.0%; Score 17; DB 4; Length 2394;
Best Local Similarity 80.0%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TCGACGAGACTTTCCTCCAAACCGA 26
||| ||||| ||||| ||||| |||||
DB 818 TCGATGGAGACTTTGCCCCACCGA 794

RESULT 10
US-10-181-660-10/c
; Sequence 10, Application US/10181660
; Patent No. 6655956
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; TITLE OF INVENTION: SUPPLEMENTED OMV VACCINE AGAINST MENINGOCOCCUS
; FILE REFERENCE: P023785W0
; CURRENT APPLICATION NUMBER: US/10/181,660
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: GB-0001067.8
; PRIOR FILING DATE: 2000-01-17
; PRIOR APPLICATION NUMBER: GB-0005699.4
; PRIOR FILING DATE: 2000-03-09
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: SeqWin99, version 1.02
; SEQ ID NO 10
; LENGTH: 2394
; TYPE: DNA
; ORGANISM: Neisseria meningitidis
US-10-181-660-10
Query Match 63.0%; Score 17; DB 4; Length 2394;
Best Local Similarity 80.0%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TCGACGAGACTTTCCTCCAAACCGA 26
||| ||||| ||||| ||||| |||||
DB 818 TCGATGGAGACTTTGCCCCACCGA 794

RESULT 11
US-08-956-171E-170
; Sequence 170, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen

```

```

; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5256
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,171E
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/009,861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Mark J. Hyman
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (240) 314-1224
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 170:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 170:
US-08-956-171E-170
Query Match 63.0%; Score 17; DB 4; Length 5333;
Best Local Similarity 80.0%; Pred. No. 76;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 CGAACGAGACTTTCCTCCAAACCGAC 27
||| ||||| ||||| ||||| |||||
DB 2397 CGACGATACCTACCTCCAAACGAC 2421

RESULT 12
US-08-701-191A-4
; Sequence 4, Application US/08701191A
; Patent No. 5942428
; GENERAL INFORMATION:
; APPLICANT: Moosa Mohammadi, Joseph Schlessinger,
; and Stevan R. Hubbard
; TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN
; OF NON-INSULIN RECEPTOR TYROSINE KINASE
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0

```

```

1 ; APPLICANT: Moosa Mohammadi, Joseph Schlessinger,
2 ; APPLICANT: and Stevan R. Hubbard
3 ; TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN
4 ; TITLE OF INVENTION: OF NON-INSULIN RECEPTOR TYROSINE KINASE
5 ; NUMBER OF SEQUENCES: 41
6 ; CORRESPONDENCE ADDRESS:
7 ; ADDRESSEE: Lyon & Lyon
8 ; STREET: 633 West Fifth Street
9 ; CITY: Los Angeles
10 ; STATE: California
11 ; COUNTRY: U.S.A.
12 ; ZIP: 90071-2066
13 ; COMPUTER READABLE FORM:
14 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
15 ; MEDIUM TYPE: storage
16 ; COMPUTER: IBM Compatible
17 ; OPERATING SYSTEM: IBM P.C. DOS 5.0
18 ; SOFTWARE: FastSeq for Windows 2.0
19 ; CURRENT APPLICATION DATA:
20 ; APPLICATION NUMBER: US/08/701,191A
21 ; FILING DATE: August 21, 1996
22 ; CLASSIFICATION: 530
23 ; PRIOR APPLICATION DATA:
24 ; APPLICATION NUMBER:
25 ; FILING DATE:
26 ; ATTORNEY/AGENT INFORMATION:
27 ; NAME: Warburg, Richard J.
28 ; REGISTRATION NUMBER: 32,327
29 ; REFERENCE/DOCKET NUMBER: 227/088
30 ; TELECOMMUNICATION INFORMATION:
31 ; TELEPHONE: (213) 489-1600
32 ; TELEFAX: (213) 955-0440
33 ; TELEX: 67-3510
34 ; INFORMATION FOR SEQ ID NO: 5:
35 ; SEQUENCE CHARACTERISTICS:
36 ; LENGTH: 1056 base pairs
37 ; TYPE: nucleic acid
38 ; STRANDEDNESS: double
39 ; TOPOLOGY: linear
40 ; MOLECULE TYPE: cDNA
41 ;
42 US-08-701-191A-5
43
44 Query Match 60.7%; Score 16.4; DB 2; Length 1056;
45 Best Local Similarity 76.9%; Pred. No. 1.1e-02;
46 Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
47
48 QY 2 TCGACGAGACTTTCGCCAACCGAC 27
49 |||||
50 DB 689 TCGACGGGACATTCACCACATCGAC 714
51 |||||
52
53 RESULT 15
54 US-09-664-526-5
55 ; Sequence 5, Application US/09664526
56 ; GENERAL INFORMATION:
57 ; Patent No. 6682921
58 ; APPLICANT: MOHAMMADI, MOOSA
59 ; APPLICANT: SCHLESSINGER, JOSEPH
60 ; APPLICANT: HUBBARD, STEVAN R.
61 ; TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN OF NON-INSULIN
62 ; TITLE OF INVENTION: RECEPTOR TYROSINE KINASES
63 ; FILE REFERENCE: 036602/0847
64 ; CURRENT APPLICATION NUMBER: US/09/664,526
65 ; CURRENT FILING DATE: 2000-09-18
66 ; PRIOR APPLICATION NUMBER: 09/188,809
67 ; PRIOR FILING DATE: 1998-11-09
68 ; PRIOR APPLICATION NUMBER: 08/701,191
69 ; PRIOR FILING DATE: 1996-08-21
70 ; NUMBER OF SEQ ID NOS: 41
71 ; SOFTWARE: PatentIn Ver. 2.1
72 ; SEQ ID NO 5
73 ; LENGTH: 1056

```

TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-664-526-5

Query Match 60.7%; Score 16.4; DB 4; Length 1056;
 Best Local Similarity 76.9%; Pred.No. 1.1e+02;
 Matches 20; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 TCGACGAGACTTCCGCCAACCGAC 27
 |||||
 Db 689 TCGACGGGACATTACCCACATCGAC 714

Search completed: April 29, 2004, 11:44:41
 Job time : 67.6471 secs

This page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

QM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 06:08:26 ; Search time 504 seconds

Title: US-10-624-714-1

Perfect score: 27

Sequence: 1 gtcgaacgagactttccccaacacgac 27

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : Published Applications_NA.*
- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
 - 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq.*
 - 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
 - 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq.*
 - 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
 - 6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq.*
 - 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq.*
 - 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq.*
 - 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq.*
 - 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq.*
 - 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq.*
 - 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq.*
 - 13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq2.*
 - 14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq.*
 - 15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq.*
 - 16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
 - 17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq.*
 - 18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq.*
 - 19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	2037	9	US-09-712-363-12
2	27	100.0	2037	13	US-10-282-122A-28150
3	27	100.0	2466	13	US-10-282-122A-26493
C 4	20.2	74.8	6862	13	US-10-221-613-37
C 5	20.2	74.8	6862	13	US-10-221-714A-21
C 6	20.2	74.8	6862	15	US-10-311-455-195
C 7	19	70.4	1728	13	US-10-282-122A-33531
C 8	18.2	67.4	497	9	US-09-783-590-5872
C 9	18	66.7	15224	13	US-10-221-613-224
C 10	17.6	65.2	1389	15	US-10-122-466A-2
C 11	17.4	64.4	332	15	US-10-060-036-4022
C 12	17.4	64.4	747	15	US-10-156-763-1654
C 13	17.4	64.4	1119	13	US-10-282-122A-18661
C 14	17.4	64.4	1364	13	US-10-424-599-64938

15	17.4	64.4	1479	16	US-10-369-493-31564
16	17.4	64.4	3312	16	US-10-369-493-27822
C 17	17.4	64.4	1230025	16	US-10-289-763-1
C 18	17.4	64.4	9025608	15	US-10-156-763-1
C 19	17.2	63.7	353	13	US-10-424-599-26166
C 20	17.2	63.7	430	16	US-10-191-803-815
C 21	17.2	63.7	514	9	US-09-884-441-251
C 22	17.2	63.7	514	10	US-09-907-969-251
C 23	17.2	63.7	514	10	US-09-827-271-251
C 24	17.2	63.7	514	15	US-10-198-053-251
C 25	17.2	63.7	1237	13	US-10-424-599-84913
C 26	17.2	63.7	16750	13	US-10-421-714A-36
C 27	17.2	63.7	16750	15	US-10-311-455-494
C 28	17.2	63.7	172637	9	US-09-805-458A-3
C 29	17.2	63.7	3673778	15	US-10-312-841-2
C 30	17	63.0	1538	13	US-10-425-114-6132
C 31	17	63.0	1640	15	US-10-192-381-27
C 32	17	63.0	1673	15	US-10-192-381-29
C 33	17	63.0	2064	13	US-10-424-599-81778
C 34	17	63.0	2127	13	US-10-282-122A-27876
C 35	17	63.0	2233	15	US-10-361-290-4
C 36	17	63.0	2235	15	US-10-361-290-5
C 37	17	63.0	2277	15	US-10-361-290-1
C 38	17	63.0	2325	13	US-10-424-599-71983
C 39	17	63.0	2391	16	US-10-320-800-5
C 40	17	63.0	2411	15	US-10-361-290-3
C 41	17	63.0	2561	13	US-10-425-114-29394
C 42	17	63.0	3542	13	US-10-424-599-96841
C 43	17	63.0	5333	8	US-08-781-986A-170
C 44	17	63.0	5333	13	US-10-329-624-170
C 45	17	63.0	6090	13	US-10-221-714A-37

ALIGNMENTS

RESULT 1

US-09-712-363-12
; Sequence 12, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712.363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 2037
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis

US-09-712-363-12

Query Match 100.0%; Score 27; DB 9; Length 2037;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGAACGAGACTTCCCAACCGAC 27
|||||
DB 1821 GTCGAACGAGACTTCCCAACCGAC 1847

RESULT 2

US-10-282-122A-28150
; Sequence 28150, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28150
; LENGTH: 2037
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis

US-10-282-122A-28150
; Sequence 28150, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28150
; LENGTH: 2037
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis

Query Match 100.0%; Score 27; DB 13; Length 2037;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGAACGAGACTTCCCAACCGAC 27
|||||
DB 1821 GTCGAACGAGACTTCCCAACCGAC 1847

RESULT 3

US-10-282-122A-26493
; Sequence 26493, Application US/10282122A
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221,613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7

US-10-282-122A-26493

US-10-282-122A-26493
; Sequence 26493, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221,613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7

Query Match 100.0%; Score 27; DB 13; Length 2466;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCGAACGAGACTTCCCAACCGAC 27
|||||
DB 2247 GTCGAACGAGACTTCCCAACCGAC 2273

RESULT 4

US-10-221-613-37/c
; Sequence 37, Application US/10221613
; Publication No. US20040029123A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
; FILE REFERENCE: 5013.1004
; CURRENT APPLICATION NUMBER: US/10/221,613
; CURRENT FILING DATE: 2002-09-13
; PRIOR APPLICATION NUMBER: PCT/EP01/02945
; DE 10013847.00
; DE 10019058.8
; DE 10019173.8
; DE 10032529.7

DE 10043826.1
PRIOR FILING DATE: 2001-03-15
2000-03-15
2000-04-06
2000-04-07
2000-06-30
2000-09-01
2000-09-30
NUMBER OF SEQ ID NOS: 428
SEQ ID NO 37
LENGTH: 6862
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-37

Query Match 74.8%; Score 20.2; DB 13; Length 6862;
Best Local Similarity 88.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGAGACTTCCCAACCGA 26
|||||
Db 6351 TCGAACGACCTTTACCAACCGA 6327

RESULT 5
US-10-221-714A-21/c
Sequence 21, Application US/10221714A
Publication No. US20040048254A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with
tumor suppressor genes and oncogenes
FILE REFERENCE: 5013.1005
CURRENT APPLICATION NUMBER: US/10/221,714A
CURRENT FILING DATE: 2003-01-21
PRIOR APPLICATION NUMBER: PCT/EP01/02955
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: DE 10013847.0
PRIOR FILING DATE: 2000-03-15
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 540
SEQ ID NO 21
LENGTH: 6862
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-714A-21

Query Match 74.8%; Score 20.2; DB 13; Length 6862;
Best Local Similarity 88.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGAGACTTCCCAACCGA 26
|||||
Db 6351 TCGAACGACCTTTACCAACCGA 6327

RESULT 6
US-10-311-455-195/c
Sequence 195, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:

APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311,455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 195
LENGTH: 6862
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-195

Query Match 74.8%; Score 20.2; DB 15; Length 6862;
Best Local Similarity 88.0%; Pred. No. 17;
Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TCGAACGAGACTTCCCAACCGA 26
|||||
Db 6351 TCGAACGACCTTTACCAACCGA 6327

RESULT 7
US-10-282-122A-33531/c
Sequence 33531, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Chlesen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 33531

; LENGTH: 1728

; TYPE: DNA

; ORGANISM: Pseudomonas syringae

US-10-282-122A-33531

Query Match

Best Local Similarity 70.4%; Score 19; DB 13; Length 1728;

Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 GTCGACGAGACTTCCCAACCGAC 27

Db 1077 GTCGACGAGACTTCTCATCCGAC 1051

RESULT 8

US-09-783-590-5872/c

; Sequence 5872, Application US/09783590

; Patent No. US20020110850A1

; GENERAL INFORMATION:

; APPLICANT: Dillon, Patrick J.

; APPLICANT: Haseltine, William A.

; APPLICANT: Li, Haodong

; APPLICANT: Rosen, Craig A.

; APPLICANT: Ruben, Steven M.

; TITLE OF INVENTION: Human Genes,

; FILE REFERENCE: PO-16.2C1

; CURRENT APPLICATION NUMBER: US/09/783,590

; PRIOR FILING DATE: 2000-02-15

; PRIOR APPLICATION NUMBER: 08/420,856

; PRIOR FILING DATE: 1995-04-12

; PRIOR APPLICATION NUMBER: 08/346,731

; NUMBER OF SEQ ID NOS: 12485

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5872

; LENGTH: 497

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (19)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (24)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (25)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (198)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (268)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (286)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (287)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (314)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (324)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (337)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (343)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (347)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (368)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (372)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (398)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (414)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (421)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (423)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (425)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (429)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (435)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (448)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (451)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (483)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (486)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (490)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (492)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (494)

; OTHER INFORMATION: n equals a,t,g, or c

; NAME/KEY: misc feature

; LOCATION: (495)

; OTHER INFORMATION: n equals a,t,g, or c

US-09-783-590-5872

Query Match 67.4%; Score 18.2; DB 9; Length 497;

Best Local Similarity 74.1%; Pred. No. 1.2e+02;

Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 GTCGACGAGACTTCCCAACCGAC 27

Db 493 GNAACAGNCITTCCTCCCAACAAAC 467

RESULT 9

US-10-221-613-224/c

; Sequence 224, Application US/10221613

; Publication No. US20040029123A1

; GENERAL INFORMATION:

APPLICANT: OLEK, Alexander
APPLICANT: PIEPENROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with Cell Cycle
FILE REFERENCE: 5013.1004
CURRENT FILING DATE: 2002-09-13
PRIOR APPLICATION NUMBER: PCT/EP01/02945
DE 10013847.00
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 1003826.1
PRIOR FILING DATE: 2001-03-15
2000-03-15
2000-04-06
2000-04-07
2000-06-30
2000-09-01
NUMBER OF SEQ ID NOS: 428
SEQ ID NO 224
LENGTH: 15224
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-221-613-224

Query Match 66.7%; Score 18; DB 13; Length 15224;
Best Local Similarity 80.8%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 TCGAAGGAGACTTCCCAACCGAC 27
|||||
DB 537 TCGAAGAAACTCCCTTAACCAAC 512
|||||

RESULT 10
US-10-122-466A-2/c
Sequence 2, Application US/10122466A
Publication No. US20030040049A1
GENERAL INFORMATION:
APPLICANT: GUEDON, ERIC
APPLICANT: ANBA-MONDOLINI, JAMILA
APPLICANT: DELORME, CHRISTINE
APPLICANT: RENAUT, PIERRE
TITLE OF INVENTION: MUTANT LACTIC BACTERIA WITH A CAPACITY FOR
OVEREXPRESSION AT LEAST ONE PEPTIDASE
FILE REFERENCE: 1096-02
CURRENT APPLICATION NUMBER: US/10/122,466A
PRIOR FILING DATE: 2002-08-27
PRIOR APPLICATION NUMBER: PCT/FR00/02869
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: FR 99/12924
PRIOR FILING DATE: 1999-10-15
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 1389
TYPE: DNA
ORGANISM: Lactococcus lactis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1389)
OTHER INFORMATION: sequence of dbpt from L. lactis IL1403
US-10-122-466A-2

Query Match 65.2%; Score 17.6; DB 15; Length 1389;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 TCGAAGGAGACTTCCCAACCGAC 25
|||||

APPLICANT: Benson, Darin R.
APPLICANT: Kalos, Michael D.
APPLICANT: Lodes, Michael J.
APPLICANT: Persing, David H.
APPLICANT: Hepler, William T.
APPLICANT: Jiang, Yugu
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
OF PANCREATIC CANCER
FILE REFERENCE: 210121.566
CURRENT APPLICATION NUMBER: US/10/060,036
CURRENT FILING DATE: 2002-01-30
NUMBER OF SEQ ID NOS: 4560
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4022
LENGTH: 332
TYPE: DNA
ORGANISM: Homo sapiens
US-10-060-036-4022

Query Match 64.4%; Score 17.4; DB 15; Length 332;
Best Local Similarity 77.8%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GTCGAAGGAGACTTCCCAACCGAC 27
|||||
DB 201 GTGGAGCGAGCTTTCTCCAAATCGGC 227
|||||

RESULT 12
US-10-156-761-1654
Sequence 1654, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1654
LENGTH: 747
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(747)
US-10-156-761-1654

Query Match 64.4%; Score 17.4; DB 15; Length 747;
Best Local Similarity 77.8%; Pred. No. 2.9e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GTCGAAGGAGACTTCCCAACCGAC 27
|||||
DB 255 GTCGCTCGCGGCTTCCCAACCGAC 281
|||||

```
RESULT 13
US-10-282-122A-18661/c
; Sequence 18661, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Haselbeck, Cheryl
; APPLICANT: Orlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,536
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18661
; LENGTH: 1119
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
US-10-282-122A-18661

Query Match          64.4%; Score 17.4; DB 13; Length 1119;
Best Local Similarity 77.8%; Pred. No. 3e+02; 6; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0;

QY 1 GTCGAACGAGACTTCCCAACCGAC 27
    |||||
DB 843 GTCGAAGAGTCTATGCCAACGCGAC 817

RESULT 14
US-10-424-599-64938/c
; Sequence 64938, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
```

```
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 64938
; LENGTH: 1364
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_29651C.1
US-10-424-599-64938

Query Match          64.4%; Score 17.4; DB 13; Length 1364;
Best Local Similarity 77.8%; Pred. No. 3e+02; 6; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0;

QY 1 GTCGAACGAGACTTCCCAACCGAC 27
    |||||
DB 477 GCGCACGTTCCITTCCTCCCAACGTC 451

RESULT 15
US-10-369-493-31564
; Sequence 31564, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31564
; LENGTH: 1479
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-31564

Query Match          64.4%; Score 17.4; DB 16; Length 1479;
Best Local Similarity 94.7%; Pred. No. 3.1e+02; 1; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 1;

QY 2 TCGAACGAGACTTTCCTCCCA 20
    |||||
DB 508 TCGAACGAGACTTTCCTCCCA 526

Search completed: April 29, 2004, 20:43:43
Job time : 512 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:23:09 ; Search time 198.118 Seconds
(without alignments)
514.627 Million cell updates/sec

Title: US-10-624-714-2

Perfect score: 24
Sequence: 1 gaccgtggctggatgacggcttc 24

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Genesecrn1980s:*
2: Genesecrn1990s:*
3: Genesecrn2000s:*
4: Genesecrn2001as:*
5: Genesecrn2001bs:*
6: Genesecrn2002s:*
7: Genesecrn2003as:*
8: Genesecrn2003bs:*
9: Genesecrn2003cs:*
10: Genesecrn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	24	100.0	2037	4	Aah51958 Mycobacte
C 2	24	100.0	2037	7	Aca40280 Prokaryot
C 3	24	100.0	2466	7	Aca38623 Prokaryot
C 4	24	100.0	110000	4	Aai199682 Mycobacte
C 5	24	100.0	110000	4	Aai199683 Mycobacte
C 6	20.8	86.7	1839	7	Aca37701 Prokaryot
C 7	18.2	75.8	960	6	Abx84464 Human cdn
C 8	18.2	75.8	960	6	Abn96954 Gene #345
C 9	18.2	75.8	1064	10	Age28348 Human KPP
C 10	18.2	75.8	1085	10	Age28349 Human KPP
C 11	18.2	75.8	1682	4	Abi19825 Drosophil
C 12	18.2	75.8	1779	4	Abi19823 Drosophil
C 13	18.2	75.8	1942	4	Abi11253 Drosophil
C 14	18.2	75.8	2169	6	Aas94762 Human DNA
C 15	18.2	75.8	14929	4	Abi11252 Drosophil
C 16	18.2	75.8	24161	4	Abi19824 Drosophil
C 17	18.2	75.8	26032	4	Abi19822 Drosophil
C 18	18.2	75.8	44861	6	Aas20000 DNA encod
C 19	18	75.0	506	8	Ach45673 Human foe
C 20	17.8	74.2	6628	4	Aak65528 Human inm
C 21	17.6	73.3	357	6	Abn22553 Human ORF
C 22	17.6	73.3	792	8	Abi10865 Alloiococ
C 23	17.6	73.3	957	7	Aca43710 Prokaryot

C 24	17.6	73.3	1212	4	ABL16213	ABL16213 Drosophil
C 25	17.6	73.3	1908	7	ACA32451	ACA32451 Prokaryot
C 26	17.6	73.3	2139	7	ACA25855	ACA25855 Prokaryot
C 27	17.6	73.3	2532	5	AAS29552	AAS29552 Human end
C 28	17.6	73.3	4871	4	ABL16212	ABL16212 Drosophil
C 29	17.6	73.3	5537	5	AAS88631	AAS88631 DNA encod
C 30	17.6	73.3	110000	8	ADBI2064_10	Continuation (11 o
C 31	17.6	73.3	110000	8	ADBI2064_11	Continuation (12 o
C 32	17.2	71.7	624	6	ABK80130	ABK80130 Bacillus
C 33	17.2	71.7	1476	6	ABS74482	ABS74482 Maize per
C 34	17.2	71.7	1690	2	AAV60818	AAV60818 Human par
C 35	17.2	71.7	1690	6	ABK91191	ABK91191 Human cdn
C 36	17.2	71.7	1956	7	ACA49216	ACA49216 Prokaryot
C 37	17.2	71.7	2352	8	AAD57506	AAD57506 Human enz
C 38	17.2	71.7	3741	7	ACA50830	ACA50830 Prokaryot
C 39	16.8	70.0	880	4	AAC99773	AAC99773 Skin cell
C 40	16.8	70.0	880	6	ABL34925	ABL34925 Murine cd
C 41	16.8	70.0	888	3	AAZ61618	AAZ61618 cDNA enco
C 42	16.8	70.0	888	4	AAC99551	AAC99551 Skin cell
C 43	16.8	70.0	888	6	ABL34703	ABL34703 Murine cd
C 44	16.8	70.0	1134	6	ABQ68052	ABQ68052 Listeria
C 45	16.8	70.0	1134	6	ABQ70042	ABQ70042 Listeria

ALIGNMENTS

RESULT 1
AAH51958/c
ID AAH51958 standard; DNA; 2037 BP.
XX
AC AAH51958;
XX
DT 04-SEP-2001 (first entry)
XX
DE Mycobacterium tuberculosis potential drug target gene SEQ ID 12.
XX
KW Drug target; growth; organism viability; characterisation; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200135317-A1.
XX
PD 17-MAY-2001.
XX
PF 13-NOV-2000; 2000WO-US0311152.
XX
PR 12-NOV-1999; 99US-0165086P.
PR 12-NOV-1999; 99US-0165124P.
PR 01-FEB-2000; 2000US-0179531P.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Eisenberg D, Rotstein SH, Marcotte EM;
XX
XX WPI; 2001-329193/34.
XX
XX P-PSDB; AAG81107.

Identifying nucleotide or polypeptide sequence for use as drug target, involves providing algorithm that analyzes a functional relationship between nucleotide or polypeptide sequences, and comparing the sequences. Disclosure; Page 55; 207pp; English.

This invention relates to a method for identifying a nucleotide or polypeptide sequence that may be a drug target, or essential for growth or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092 represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium tuberculosis proteins which are potential drug targets. The DNA and protein sequences are used to illustrate the method of the invention. The method involves providing an unknown nucleotide or polypeptide sequences, and comparing it to a number of sequences along with at least one algorithm capable of analysing a functional relationship between

CC nucleotide and polypeptide sequences. The method is useful for
 CC characterizing the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism
 XX
 SQ Sequence 2037 BP; 396 A; 639 C; 651 G; 351 T; 0 U; 0 Other;
 Query Match 100.0%; Score 24; DB 4; Length 2037;
 Best Local Similarity 100.0%; Pred. No. 0.52;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 GACCGTGGCTGGATGACGGTCTC 24
 |||||
 Db 1932 GACCGTGGCTGGATGACGGTCTC 1909
 RESULT 2
 ACA40280/c
 ID ACA40280 standard; DNA; 2037 BP.
 XX
 AC ACA40280;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Prokaryotic essential gene #21937.
 XX
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US0009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR P-PSDB; ABU36410.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 14; SEQ ID NO 28150; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an

CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2037 BP; 396 A; 639 C; 651 G; 351 T; 0 U; 0 Other;
 Query Match 100.0%; Score 24; DB 7; Length 2037;
 Best Local Similarity 100.0%; Pred. No. 0.52;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 QY 1 GACCGTGGCTGGATGACGGTCTC 24
 |||||
 Db 1932 GACCGTGGCTGGATGACGGTCTC 1909
 RESULT 3
 ACA38623/c
 ID ACA38623 standard; DNA; 2466 BP.
 XX
 AC ACA38623;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Prokaryotic essential gene #20280.
 XX
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 XX
 OS Mycobacterium bovis.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US0009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR P-PSDB; ABU34753.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 14; SEQ ID NO 26493; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 2466 BP; 454 A; 808 C; 800 G; 404 T; 0 U; 0 Other;

Query Match 100.0%; Score 24; DB 7; Length 2466;
Best Local Similarity 100.0%; Pred. No. 0.53;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACCGTGGCTGGATGACGGTCTC 24
DB 2364 GACCGTGGCTGGATGACGGTCTC 2341

RESULT 4

WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682

Fragment Name	Begin	End
AAI99682_00	1	110000
AAI99682_01	100001	210000
AAI99682_02	200001	310000
AAI99682_03	300001	410000
AAI99682_04	400001	510000
AAI99682_05	500001	610000
AAI99682_06	600001	710000
AAI99682_07	700001	810000
AAI99682_08	800001	910000
AAI99682_09	900001	1010000
AAI99682_10	1000001	1110000
AAI99682_11	1100001	1210000
AAI99682_12	1200001	1310000
AAI99682_13	1300001	1410000
AAI99682_14	1400001	1510000
AAI99682_15	1500001	1610000
AAI99682_16	1600001	1710000
AAI99682_17	1700001	1810000
AAI99682_18	1800001	1910000
AAI99682_19	1900001	2010000
AAI99682_20	2000001	2110000
AAI99682_21	2100001	2210000
AAI99682_22	2200001	2310000
AAI99682_23	2300001	2410000
AAI99682_24	2400001	2510000
AAI99682_25	2500001	2610000
AAI99682_26	2600001	2710000
AAI99682_27	2700001	2810000
AAI99682_28	2800001	2910000
AAI99682_29	2900001	3010000
AAI99682_30	3000001	3110000

WP AAI99682_31 3100001 3210000
WP AAI99682_32 3200001 3310000
WP AAI99682_33 3300001 3410000
WP AAI99682_34 3400001 3510000
WP AAI99682_35 3500001 3610000
WP AAI99682_36 3600001 3710000
WP AAI99682_37 3700001 3810000
WP AAI99682_38 3800001 3910000
WP AAI99682_39 3900001 4010000
WP AAI99682_40 4000001 4110000
WP AAI99682_41 4100001 4210000
WP AAI99682_42 4200001 4310000
WP AAI99682_43 4300001 4410000
WP AAI99682_44 4400001 4411529
ID AAI99682 standard; DNA; 4411529 BP.
XX
AC AAI99682;
DT 15-JAN-2002 (first entry)
XX
XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.
DE Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
XX variation; epidemiology; patient treatment; epidemic monitoring; ds.
KW Mycobacterium tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN US294328-B1.
XX
PD 25-SEP-2001.
XX
PF 24-JUN-1998; 98US-00103840.
XX
PR 24-JUN-1998; 98US-00103840.
XX
XX (GENO-) INST GENOMIC RES.
XX
PI Fleischmann RD, White OR, Fraser CM, Venter JC;
XX
DR WPI; 2001-647261/74.
XX
PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
PT determining the nucleotide sequence of the strain at positions in the
PT genome corresponding to positions where *M. tuberculosis* strains CDC 1551
PT and H37Rv differ.
XX
PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.
XX
CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of *M. tuberculosis* strains CDC 1551 (AAI99683) and
CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC *M. tuberculosis* and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from USPTO at seqdata.uspto.gov/sequence.html?docID=6294328B1
XX
SQ Sequence 4411529 BP; 758555A; 1449983C; 1444602G; 758379T; 0U; 0Other;
Query Match 100.0%; Score 24; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACCGTGGCTGGATGACGGTCTC 24

DB 55592 GACCGTGGCTGGATGACGGTCTC 55569

RESULT 5

AAI99683 00/c
 WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683
 WP Fragment Name Begin End
 WP AAI99683_00 1 110000
 WP AAI99683_01 100001 210000
 WP AAI99683_02 200001 310000
 WP AAI99683_03 300001 410000
 WP AAI99683_04 400001 510000
 WP AAI99683_05 500001 610000
 WP AAI99683_06 600001 710000
 WP AAI99683_07 700001 810000
 WP AAI99683_08 800001 910000
 WP AAI99683_09 900001 1010000
 WP AAI99683_10 1000001 1110000
 WP AAI99683_11 1100001 1210000
 WP AAI99683_12 1200001 1310000
 WP AAI99683_13 1300001 1410000
 WP AAI99683_14 1400001 1510000
 WP AAI99683_15 1500001 1610000
 WP AAI99683_16 1600001 1710000
 WP AAI99683_17 1700001 1810000
 WP AAI99683_18 1800001 1910000
 WP AAI99683_19 1900001 2010000
 WP AAI99683_20 2000001 2110000
 WP AAI99683_21 2100001 2210000
 WP AAI99683_22 2200001 2310000
 WP AAI99683_23 2300001 2410000
 WP AAI99683_24 2400001 2510000
 WP AAI99683_25 2500001 2610000
 WP AAI99683_26 2600001 2710000
 WP AAI99683_27 2700001 2810000
 WP AAI99683_28 2800001 2910000
 WP AAI99683_29 2900001 3010000
 WP AAI99683_30 3000001 3110000
 WP AAI99683_31 3100001 3210000
 WP AAI99683_32 3200001 3310000
 WP AAI99683_33 3300001 3410000
 WP AAI99683_34 3400001 3510000
 WP AAI99683_35 3500001 3610000
 WP AAI99683_36 3600001 3710000
 WP AAI99683_37 3700001 3810000
 WP AAI99683_38 3800001 3910000
 WP AAI99683_39 3900001 4010000
 WP AAI99683_40 4000001 4110000
 WP AAI99683_41 4100001 4210000
 WP AAI99683_42 4200001 4310000
 WP AAI99683_43 4300001 4403765
 ID AAI99683 standard; DNA; 4403765 BP.
 XX AC AAI99683;
 XX PF 15-JAN-2002 (first entry)
 XX DT
 XX DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
 XX KW Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
 KW variation; epidemiology; patient treatment; epidemic monitoring; ds.
 XX OS Mycobacterium tuberculosis.
 XX PN US6294328-B1.
 XX PD 25-SEP-2001.
 XX PF 24-JUN-1998; 98US-00103840.
 XX PR 24-JUN-1998; 98US-00103840.
 XX PA (GENO-) INST GENOMIC RES.
 XX PI Fleischmann RD, White OR, Fraser CM, Venter JC;
 XX WPI; 2001-647261/74.
 XX DR

XX PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
 PT determining the nucleotide sequence of the strain at positions in the
 PT genome corresponding to positions where M. tuberculosis strains CDC 1551
 PT and H37Rv differ.
 XX PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
 XX CC The invention relates to evaluating strain variation within and between
 CC different populations of the tuberculosis bacterial pathogen,
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the
 CC nucleotide sequence of the first strain at positions in the complete
 CC sequence of the genome that correspond to positions that differ in the
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AAI99683) and
 CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
 CC M. tuberculosis and has valuable application in the fields of
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic
 CC monitoring. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from USPTO at seqdata.uspto.gov/sequence.html?DocID=6294328B1
 XX SQ Sequence 4403765 BP; 757105A; 1447799C; 1441301G; 757371T; 0U; 1890Other;
 Query Match 100.0%; Score 24; DB 4; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 0.66;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GACCGTGGCTGGATGACGGTCTC 24
 Db 55534 GACCGTGGCTGGATGACGGTCTC 55511
 RESULT 6
 ACA37701/c
 ID ACA37701 standard; DNA; 1839 BP.
 XX AC ACA37701;
 XX DT 19-JUN-2003 (first entry)
 XX DE Prokaryotic essential gene #19359.
 XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 XX OS Mycobacterium avium.
 XX PN WO200277193-A2.
 XX PD 03-OCT-2002.
 XX PF 21-MAR-2002; 2002WO-US009107.
 XX PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX PA (ELIT-) ELITRA PHARM INC.
 XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
 XX WPI; 2003-029926/02.
 DR P-PSDB; ABU33831.
 XX PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX PS Claim 14; SEQ ID NO 25571; 1766pp; English.
 XX DR

XX Homo sapiens.

XX ADE28349;
 XX 29-JAN-2004 (first entry)
 DT Human KPP cDNA - SEQ ID 60.
 XX
 XX kinase; phosphatase; KPP; hepatotropic; antiarteriosclerotic;
 KW antipsoriatic; cytoskeletal; haemostatic; muscular; cerebroprotective;
 KW nontropic; ophthalmological; anticonvulsant; vasotropic; neuroprotective;
 KW antiparkinsonian; antiasthmatic; antianaemic; antiasthmatic;
 KW antidiabetic; antiinflammatory; osteopathic; antiarthritic;
 KW antirheumatic; dermatological; virucide; antibacterial; fungicide;
 KW antiparasitic; protozoacide; antihelminthic; antigout; cardiovascular;
 KW antiarteriosclerotic; immunosuppressive; cell proliferative; cirrhosis;
 KW hepatitis; arteriosclerosis; psoriasis; primary thrombocytopenia; cancer;
 KW developmental; renal tubular acidosis; Becker's muscular dystrophy;
 KW gonadal dysgenesis; hypothyroidism; seizure; neurological;
 KW Pick's disease; cataract; epilepsy; ischaemic cerebrovascular; stroke;
 KW Alzheimer's; Parkinson's; dementia; autoimmune; inflammatory; AIDS;
 KW allergy; anaemia; asthma; diabetes mellitus; bronchitis; osteoporosis;
 KW osteoarthritis; rheumatoid arthritis; contact dermatitis; gout;
 KW lipid disorder; cholestasis; Gaucher's; diabetes; atherosclerosis; liver;
 KW viral; bacterial; fungal; parasitic; protozoan; helminthic infection;
 KW trauma; gene therapy; human; ss; gene.
 XX
 OS Homo sapiens.
 XX
 XX WO2003080805-A2.
 XX
 PD 02-OCT-2003.
 XX
 XX 18-MAR-2003; 2003WO-US008715.
 XX
 XX 19-MAR-2002; 2002US-0366088P.
 PR 29-MAR-2002; 2002US-0369248P.
 XX
 XX (INCY-) INCYTE CORP.
 PA
 XX Chien D, Jin P, Hawkins PR, Baughn VR, Becha SD, Chang H;
 PI Ding L, Elliott VS, Emerling BM, Gandhi AR, Gietzen KJ, Griffin JA;
 PI Gururajan R, Hafalia AJA, Ison CH, Kable AE, Khare R, Lee SY;
 PI Lee BA, Lu Y, Marquis JP, Lehr-Mason PM, Ramkumar J, Richardson TW;
 PI Swarnakar A, Tran UK, Chawla NK, Yao MG, Yue H, Bhatia U;
 PI Burrill JD, Lee S, Blake JJ, Ho A, Zheng W;
 XX WPI; 2004-011523/01.
 DR P-PSDB; ADE28297.
 DR
 XX New human kinases and phosphatases, and polynucleotides encoding them,
 PT useful for treating, preventing or diagnosing e.g. cell proliferative
 PT disorders, inflammatory, autoimmune, viral, bacterial, parasitic or
 PT fungal diseases.
 XX
 PS Claim 5; SEQ ID NO 60; 340pp; English.
 XX
 XX The invention relates to a novel isolated kinase and phosphatase (KPP)
 CC polypeptide. The polypeptide of the invention demonstrates hepatotropic,
 CC antiarteriosclerotic, antipsoriatic, cytoskeletal, haemostatic, muscular,
 CC cerebroprotective, nontropic, ophthalmological, anticonvulsant,
 CC antianaemic, neuroprotective, antiparkinsonian, antiasthmatic,
 CC antirheumatic, antiasthmatic, antidiabetic, antiinflammatory, osteopathic,
 CC antiarthritic, antirheumatic, dermatological, virucide, antitubercular,
 CC fungicide, antiparasitic, protozoacide, antihelminthic, antigout,
 CC cardiovascular, antiarteriosclerotic and immunosuppressive activities.
 CC The KPP polypeptides may be useful for diagnosing, treating or preventing
 CC cell proliferative disorders including cirrhosis, hepatitis,
 CC arteriosclerosis, psoriasis, primary thrombocytopenia and cancer,
 CC developmental disorders such as renal tubular acidosis, Becker's muscular
 CC dystrophy, gonadal dysgenesis, hypothyroidism or seizures, neurological
 CC disorders e.g. Pick's disease, cataract, epilepsy, ischaemic
 CC cerebrovascular disease, stroke, Alzheimer's disease, Parkinson's disease
 CC or dementia, autoimmune or inflammatory disorders including AIDS,

CC allergies, anaemia, asthma, diabetes mellitus, bronchitis, osteoporosis,
 CC osteoarthritis, rheumatoid arthritis, contact dermatitis or gout and
 CC lipid disorders such as cholestasis, Gaucher's disease, diabetes,
 CC atherosclerosis or liver disease, as well as viral, bacterial, fungal,
 CC parasitic, protozoan or helminthic infections and trauma. Furthermore,
 CC the polypeptide may be utilised during gene therapy procedures. The
 CC current sequence is that of the human KPP cDNA of the invention.
 XX
 SQ Sequence 1085 BP; 262 A; 295 C; 328 G; 200 T; 0 U; 0 Other;
 Query Match 75.8%; Score 18.2; DB 10; Length 1085;
 Best Local Similarity 87.0%; Pred. No. 1.9e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 ACCGTGGCTGGATGACGGTCTC 24
 DB 835 ACCGTGGCTGGATGACGGTCTC 813
 RESULT 11
 ABL19825/c
 ID ABL19825 standard; DNA; 1682 BP.
 XX
 AC ABL19825;
 DT 26-MAR-2002 (first entry)
 XX
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 10948.
 DE
 XX Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 KW
 OS Drosophila melanogaster.
 XX
 XX WO200171042-A2.
 PN
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US009231.
 PF
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 PT
 XX Claim 1; SEQ ID NO 10948; 21pp + Sequence Listing; English.
 PS
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1682 BP; 440 A; 511 C; 436 G; 295 T; 0 U; 0 Other;
 Query Match 75.8%; Score 18.2; DB 4; Length 1682;
 Best Local Similarity 87.0%; Pred. No. 1.9e+02;
 Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GACCGTGGCTGGATGACGGTCTC 23

PI Shiffman D, Somogyi R, Lawn R, Seilhamer JJ, Porter GJ, Mikita T;
PI Tai J;
XX
DR WPI; 2002-010925/01.
XX
XX Composition useful for diagnosis of conditions, disorders or diseases
PT associated with atherosclerosis, comprises several polynucleotides that
PT are differentially expressed in foam cell development.
XX
XX Claim 1; Page 75-76; 315pp; English.
XX
XX The present invention relates to the isolation of human polynucleotide
CC sequences that are differentially expressed during foam cell
CC differentiation. The polynucleotide sequences of the invention or a
CC composition comprising these polynucleotides are useful as a high
CC throughput method for detecting altered expression of one or more
CC polynucleotides in a sample. The polynucleotides can be used in the
CC diagnosis of disorders associated with foam cell development such as
CC atherosclerosis, cerebral stroke, and cardiovascular disorders such as
CC coronary artery disease. The polynucleotide sequences can also be used as
CC PCR primers and probes. The polynucleotides of the invention are also
CC useful in gene therapy. AAS94746-AAS95021 represent the human
CC polynucleotide sequences of the invention which are differentially
CC expressed during foam cell differentiation
XX
XX Sequence 2169 BP; 426 A; 615 C; 679 G; 449 T; 0 U; 0 Other;
Query Match 75.8%; Score 18.2; DB 6; Length 2169;
Best Local Similarity 87.0%; Pred. No. 1.9e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 ACCGTGGGTGGATGACGGTCTC 24
DB 1092 ACCGTGGCTGGACGACGATCTC 1070
RESULT 15
ABL11252
ID ABL11252 standard; cDNA; 14929 BP.
AC ABL11252;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 28238.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
PR
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
DR P-PSDB; ABB67149.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 28238; 21pp + Sequence Listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 14929 BP; 4248 A; 3273 C; 3204 G; 4204 T; 0 U; 0 Other;
Query Match 75.8%; Score 18.2; DB 4; Length 14929;
Best Local Similarity 87.0%; Pred. No. 2.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GACCGTGGGTGGATGACGGTCT 23
DB 12264 GCGCGTGGTCTGGATGACGATCT 12286
Search completed: April 29, 2004, 04:56:52
Job time : 201.218 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:25:04 ; Search time 362.165 Seconds
(without alignments)
2872.264 Million cell updates/sec

Title: US-10-624-714-2

Perfect score: 24

Sequence: 1 gaccgtggctggatgacggtctc 24

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pin.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	24	100.0	16644	1	AE006918 Mycobacte
C 2	24	100.0	38721	1	AD000013 Mycobacte
C 3	24	100.0	341957	15	BX842572 Mycobacte
C 4	24	100.0	343050	1	BX248334 Mycobacte
C 5	19.2	80.0	2537	8	AK106569 Oryza sat
C 6	19.2	80.0	67000	8	AP005251 Oryza sat
C 7	13.2	80.0	71195	2	AP003893 Oryza sat
C 8	18.4	76.7	5749	1	SCWHIE
C 9	18.4	76.7	300100	1	SC0939123
C 10	18.2	75.8	939	12	AY335636
C 11	18.2	75.8	960	6	AX410805
C 12	18.2	75.8	960	9	HSU89606
C 13	18.2	75.8	995	4	AF125374
C 14	18.2	75.8	1069	9	AY303972
C 15	18.2	75.8	1227	9	BC005825
C 16	18.2	75.8	1331	9	BC000123
C 17	18.2	75.8	2169	6	AX281608
C 18	18.2	75.8	2789	3	BT003579
C 19	18.2	75.8	10029	1	AE012348
C 20	18.2	75.8	10507	1	AE012250
C 21	18.2	75.8	11078	1	AE011786
C 22	18.2	75.8	11188	1	AE006001
C 23	18.2	75.8	62198	2	AC013832
C 24	18.2	75.8	110000	2	AC003656
C 25	18.2	75.8	120116	8	AC092390
C 26	18.2	75.8	142956	8	AP004641
C 27	18.2	75.8	145540	9	AP001052
C 28	18.2	75.8	162770	3	AC008309
C 29	18.2	75.8	165701	8	AP004357
C 30	18.2	75.8	168745	3	AC016132
C 31	18.2	75.8	168745	2	AC119552
C 32	18.2	75.8	175992	2	AC123897
C 33	18.2	75.8	197925	2	AC091308
C 34	18.2	75.8	204617	2	BX890637
C 35	18.2	75.8	211918	10	AL928868
C 36	18.2	75.8	216059	2	AC111899
C 37	18.2	75.8	220218	9	AC006288
C 38	18.2	75.8	220336	2	AC020256
C 39	18.2	75.8	232241	10	AC104326
C 40	18.2	75.8	234985	2	AC139570
C 41	18.2	75.8	241480	3	AE003734
C 42	18.2	75.8	248329	2	AC109989
C 43	18.2	75.8	249871	2	AC118143
C 44	18.2	75.8	266661	2	AC126523
C 45	18.2	75.8	277239	2	AC123077

ALIGNMENTS

RESULT 1
AE006918/c
LOCUS AE006918 16644 bp DNA linear BCT 27-APR-2001
DEFINITION Mycobacterium tuberculosis CDC1551, section 4 of 280 of the complete genome.
ACCESSION AE006918 AE000516
VERSION AE006918.1 GI:13879090
KEYWORDS
SOURCE Mycobacterium tuberculosis CDC1551
ORGANISM Mycobacterium tuberculosis CDC1551
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacteriineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 16644)
AUTHORS Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,

Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J. F., Nelson, W. C., Umayam, L. A., Emolaeva, M., Salzberg, S. L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains

JOURNAL REFERENCE

2 (bases 1 to 16644)
Fleischmann, R. D., Alland, D., Eisen, J. A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J. F., Nelson, W. C., Umayam, L. A., Emolaeva, M., Salzberg, S. L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES

Location/Qualifiers
1..16644

/organism="Mycobacterium tuberculosis CDC1551"

/mol_type="genomic DNA"

/strain="CDC1551"

/db_xref="taxon:83331"

/note="clinical strain"

complement(86..712)

/gene="MT0048"

complement(86..712)

/gene="MT0048"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

complement(871..1605)

/gene="MT0049"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=1

/product="hypothetical protein"

/protein_id="AAK44273.1"

/db_xref="GI:13879094"

/translation="MLSDELTLGLDEFAALLAENAEQAVNGPLPEVERVOAGISAUR

WGSAPRVIFLHGCGNAHTWDTVIVGLGSPALAVDLFGHGSAWREDGNYSPQNSE

TLAPVLELACAEFVVGMSLGLTALRAAPADLVGLVLDVDTPSALQRAELTA

EOGTVALMHGEREPFSPQMLDITIAAAPHRDVKLSRRGVENSRRLDNGNWRVD

AIRFEGPAGLMDVDALSAPIITLVRCSSGFSVTDQTAELHRRATHFRGHIHVKSG

HSVQDQFALIEIVRGVLDTR"

complement(3526..4629)

/gene="MT0052"

complement(3526..4629)

/gene="MT0052"

/note="similar to GP:4808395; identified by sequence

similarity; putative"

/codon_start=1

/transl_table=1

/product="1L-myo-inositol-1-phosphate synthase"

/protein_id="AAK44274.1"

/db_xref="GI:13879095"

/translation="MSEHQSLEPAPEASTEVRVAIVGVNCASILVQGVVYVYVADDT

TVPGLMHVRGPHVHRDVKFAAFDVAQKVGPDLSDAIFASENNTIKIADVPTVVI

VQRTLDGIGKYADTIELSDAPVDVQVQALKEAKVDVVLVSLIPVSGEADKFAQC

AIDAGVAFNALPVFIASDPVAKKFTDAGVPIVDDIKSQVGATITHRVLAKLFEDR

GVQDRTQNLNVGNMDFLNLERLEKISKTKTQVNTSKREFTKDVIHIGPSDH

VGWLDRKWAYVRLGFAFGVDPLNLEYKLEVMDSFNSAGVIIDAVRAAKIAKDRGTG

GPVTPASAYLMKSPPEQLPDDIARAQLEEEFIIG"

complement(4650..5232)

/gene="MT0053"

complement(4650..5232)

/gene="MT0053"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=1

/product="conserved hypothetical protein"

/protein_id="AAK44275.1"

/db_xref="GI:13879096"

/translation="MLELAILGLIESPMHGYELRKLTVGLGAFRAFSVGSLYPALR

RMQADGLIAENAPAGTVFRRARRVYQITDKRRRFGELVADTGPVNTDQFGVHLA

FNNTPAEARMRLGRRRQVEERREGLREAVARASFDRTYRQLHQLGLESEREV

KWNLITAEARAEAPAEQT"

complement(5333..6202)

/gene="MT0054"

complement(5333..6202)

/gene="MT0054"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=1

/product="hypothetical protein"

/protein_id="AAK44276.1"

/db_xref="GI:13879097"

/translation="MAKWLGAPLARGVSTRAKDSRDQACRILDALDGLSMBEE

HRERSAATKAVTIGDQLRVLDQVESAQPAQPAKRAKTELGLAAFAFVAVSL

GVGIGWGYNTSRPLDFTDPGAKPDGIAPIVLTTPRQLHSLGLGLTLEQTRKRF

DTMGVRLVIVPEVALDRVDPADRRRLVAVTYRGWDATSSAKSIADSVVDLSKFD

AKTAGIMRGAPETLGLKQSDVKSMYLIIVDPADPTTPAALSLSLYVSSDYGGGILVF

AGDGTIKHVSYPF"

6336..6749

/gene="MT0055"

6336..6749

/gene="MT0055"

/note="identified by Glimmer2; putative"

/codon_start=1

/transl_table=1

/product="conserved hypothetical protein"

/protein_id="AAK44277.1"

/db_xref="GI:13879098"

/translation="MDYTLRRSLAEVYSGRTGVSEVCDANPYLLRAAKFHGKPSRV

ICPICRKEOLTLSVWVFGHGLGAVSGSARTAEELILATRFSEFAVHVVEVCRTCSWN

HLVKSIVLGAARPARPRGSGGTRTAENGARTASE"

```
gene
6742. .9204
/genes="MT0056"
CDS
6742. .9204
/genes="MT0056"
/note="similar to PID.886307; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="penicillin-binding protein"
/protein_id="AAK4278.1"
/db_xref="GI:13879099"
/translation="MNSDGRHHQSSGAPRGANPQRCQVPPDDELTAIPVPTDDR
SAPHDSIRKALGAPNPPPDLEVTALAAPEKPPRQDQGLGRRPPGP
GPGSGGAGRLPQPRDLPRVQGNWIRRSYLTAAVILLPMTVTFWYLIIVD
VPKGDRTNQVSTILASGSEAKIIPPEGRVDNLQVPMHVQRAVIAEDRNFY
SNPFGFTFAZAKNNLFGDLQGGSTITQQYVKNALQVSHQSGMLRKAKELVI
ATKMSGKWDKQVLAQYNIIVFGAGYGIISAARAYFKPVEQLTVAEGALLALIR
RSTLDPDPEGAHARMWVLDGMVETKALSPNDRAAOVFPETVPDLARAENOTKG
PUGLIEROTVRELLELFDNIDCOLNTQGLAVTTIDPOARAARAKAVAKYLDGQPD
RAVSDIHNGNAVAYGDDNANGDFPAQAGLQTSSEFKFALVALEQGLGLGVQV
DPSPLVDGKITNVBEGCGTNTAELKMSLNTSYRMLKNGPQPOADAHAQA
GIASSPGVAHTLSEDKNTGPNNGIVLQYQYTRVIDMASAYATLAAAGIYHPHFVQ
KVVSAHQVLFDASTDNTGDDRIKRAVDNVTAAEPIAGYSRGNLAGSPDSIAKT
GTTQFGDTTANKDAMWVGTPLSLTAVWGTGVKDEPLVTAGSAAIYSGLPSDIWKA
TYDGLKGTSTNETFPRKPTVGVGAGVPPPPPEPVPSETVIQPTVEIAPGITPIGP
PTITLAPPPAPPAATPTPPP"
9246. .10883
/genes="MT0057"
CDS
9246. .10883
/genes="MT0057"
/note="identified by Glimmer2; putative"
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK4279.1"
/db_xref="GI:13879100"
/translation="MAADLRSDNRDCPSRTDVLGAALANVVGVPGRHALICRTRLM
TPLRVFAIALFALGWSKCAACLOSTGTGPGDORVANDNORAYVOLCYSDTVPVLY
GAELLSGPFKPVKSWIETSGNCTPOLRYDGLAVRYMEYVPLTGYOVLWIAKTY
TALSKVAPLVAEVNMFNVAAPGLALAWLTITWATSLAGRRINDAALVAASPLVI
FQITFDALATGLATSGLLAWRRFVLAVGLIGLSAAKULPLFLYPLLLGLGIRA

Query Match 100.0%; Score 24; DB 1; Length 16644;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACCGTGGCTGGATGACGGTCTC 24
|||||
DB 9099 GACCGTGGCTGGATGACGGTCTC 9076

RESULT 2
MSGY219 38721 bp DNA linear BCT 10-DEC-1996
LOCUS Mycobacterium tuberculosis sequence from clone y219.
DEFINITION AD000013
ACCESSION AD000013
VERSION AD000013.1 GI:1717736
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE 1 (bases 1 to 38721)
AUTHORS Du, L.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
Beaver Street, Waltham, MA, USA, 02154 du@cric.com
COMMENT GSDS:S:1004710
FEATURES
source
1. .38721
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"

/db_xref="taxon:1773"
/clone="y219"

ORIGIN
Query Match 100.0%; Score 24; DB 1; Length 38721;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACCGTGGCTGGATGACGGTCTC 24
|||||
DB 6132 GACCGTGGCTGGATGACGGTCTC 6155

RESULT 3
BX842572/c
ID BX842572 standard; circular genomic DNA; PRO; 341957 BP.
XX BX842572; AL021427; AL021428; AL021926; AL021927; AL021928; AL021929;
AC AL021930; AL123456; Z74410; Z80233; Z80775; Z86089; Z92669; Z92770; Z96071;
AC Z97050;
XX BX842572.1
SV
XX
XX
DT 21-NOV-2003 (Rel. 77, Created)
DT 21-NOV-2003 (Rel. 77, Last updated, Version 1)
XX
XX Mycobacterium tuberculosis H37Rv complete genome; segment 1/13
DE complete genome.
XX
XX Mycobacterium tuberculosis H37Rv
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
OC Mycobacterium tuberculosis complex; Mycobacterium tuberculosis.
XX
XX [1]
RN MEDLINE; 98295987.
RX PUBMED; 9634230.
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eigmeier K., Gas S., Barry III C.E., Tekala F., Badcock K.,
Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K.,
Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagsis K.,
Krogh A., McLean J., Moule S., Murphy J., Oliver S., Osborne J.,
Quail M.A., Rajandram M.A., Rogers J., Rutter S., Seeger K., Skelton S.,
Squires S., Scars R., Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
"Deciphering the biology of Mycobacterium tuberculosis from the complete
genome sequence";
RL Nature 393:537-544 (1998).
XX
XX [2]
RX PUBMED; 12368430.
RX Camus J.C., Pryor M.J., Medigue C., Cole S.T.;
"Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv";
RL Microbiology 148:2967-2973 (2002).
XX
XX [3]
RX 1-341957
RA Parkhill J.;
RT
RL Submitted (11-JUN-1998) to the EMBL/GenBank/DDJ databases.
RL Submitted on behalf of the Mycobacterium tuberculosis sequencing and
mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut
Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail:
parkhill@sanger.ac.uk
XX
XX Notes:
CC Details of M. tuberculosis sequencing at the Sanger Centre
CC are available on the World Wide Web.
CC (URL, http://www.sanger.ac.uk/Projects/M\_tuberculosis/)
XX
XX Key Location/Qualifiers
FH
```

FT 1..341957
 FT /db_xref="taxon:83332"
 FT /mol_type="genomic DNA"
 FT /organism="Mycobacterium tuberculosis H37Rv"
 FT /strain="H37Rv"
 FT 1..1524
 FT CDS
 FT /evidence=EXPERIMENTAL
 FT /note="Rv0001, (MT0001, MV029.01, P49993), len: 507 aa.
 FT dnaA, chromosomal replication initiator protein (see
 FT citations below), equivalent to other Mycobacterial
 FT CHROMOSOMAL REPLICATION INITIATOR PROTEINS e.g.
 FT P46388 [DnaA_MYCE from Mycobacterium leprae (502 aa);
 FT Q9L7L7 [DnaA_MYCPA from Mycobacterium paratuberculosis (509
 FT aa); P49590 [DnaA_MYCAV from Mycobacterium avium (508 aa);
 FT P49992 [DnaA_MYCSM from Mycobacterium smegmatis (504 aa);
 FT etc. Also highly similar to others except in N-terminus
 FT e.g. Q92H75 [DnaA_STRCH CHROMOSOMAL REPLICATION INITIATOR
 FT PROTEIN from Streptomyces chrysomallus (624 aa);
 FT Q92H76 [DnaA_STRRE from Streptomyces reticuli (643 aa);
 FT DnaA_ECOLI|P03004|B3702 chromosomal replication initiator
 FT protein from Escherichia coli strain K12 (467 aa), FASTA
 FT scores: opt: 986, E(): 0, (43.2% identity in 389 aa
 FT overlap); etc. Contains P500017 ATP/GTP-binding site motif
 FT A (P-loop) and P501008 DnaA protein signature. BELONGS TO
 FT THE DnaA FAMILY. Note that the first base of this gene has
 FT been taken as base 1 of the Mycobacterium tuberculosis
 FT H37Rv genomic sequence."
 FT /transl_table=11
 FT /function="PLAYS AN IMPORTANT ROLE IN THE INITIATION AND
 FT REGULATION OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN
 FT OF REPLICATION; IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA
 FT AT A 9 BP CONSENSUS (DnaA BOX): 5'-TTATC(C/A)(C/A)A-3'.
 FT DnaA BINDS TO ATP AND TO ACIDIC PHOSPHOLIPIDS. DnaA PROTEIN
 FT BINDS THE ORIGIN OF REPLICATION (oric), ATP AND ADP, AND
 FT EXHIBITS WEAK ATPase ACTIVITY."
 FT /gene="dnaA"
 FT /locus_tag="RV0001"
 FT /product="CHROMOSOMAL REPLICATION INITIATOR PROTEIN DnaA"
 FT /translation="MTDDPSGFTVYNNVAVSLGDKPKVDGDSNDLSAPLPQOR
 FT AWNLVQPIIVGFPALLSPVPSFVGNIEIRAPITDALSRRHQIQLGVRIAPA
 FT TDADDTTPSPNPATSPDTITDDEIDSSAARGDNQSWSPSYFTFPHNTDSATA
 FT GVTSLNRRYFTVTCAGNFEAAALAAIAEPAPAYNPLFIWGESLGKTHLHAAG
 FT NYAQLFPGRVYVTEFTNDINSLDRDKVAFKRSYRDVLLVDIIQIFIRKGG
 FT EGFTEFTLNLNANKQIVISDRPKQLATLEDRLRVFEWGLITVQPPLETRIA
 FT ILKKAQMERLAVDDVLELIASSIRNTELEGALIRVTFASLKNTPIDKALAEIVL
 FT RDLADANQMOISAATIMATAEYFTTVEELRGPKTPALAOQSQILAMYLCREITDLS
 FT LPKIGAFGRDHTVYVYAKKILSEMAERREVFVHVKEITTRIGRSK"
 FT 622..645
 FT /note="P500017 ATP/GTP-binding site motif A (P-loop)"
 FT /notes="P501008 DnaA protein signature"
 FT 1384..1440
 FT CDS
 FT 2052..3260
 FT /evidence=EXPERIMENTAL
 FT /note="Rv0002, (MTV029.02, MTCY10H4.0), len: 402 aa. dnaN,
 FT DNA polymerase III (beta chain) (EC 2.7.7) (see citations
 FT below), equivalent to other Mycobacterial DNA POLYMERASES
 FT III BETA CHAIN e.g. NP 301130.1|NC 002677 from
 FT Mycobacterium leprae (399 aa); Q9L7L6|Dp3B_MYCPA from
 FT Mycobacterium avium subsp. paratuberculosis (399 aa);
 FT P52851|Dp3B_MYCSM from Mycobacterium smegmatis (397 aa);
 FT etc. Also highly similar to others e.g. P27903|Dp3B_STRCO
 FT DNA POLYMERASE III BETA CHAIN from Streptococcus coelicolor
 FT (376 aa), FASTA scores: opt: 1189, E(): 0, (52.8% identity
 FT in 337 aa overlap); P21174|Dp3B_MICLU from Micrococcus
 FT luteus (310 aa); P52023|Dp3B_SYNP7 from Synechococcus sp.
 FT strain PC 7942 (375 aa); etc. Overlaps and extends CDS in
 FT neighbouring cosmid MTCY10H4.01."
 FT /transl_table=11
 FT /EC_number="2.7.7."
 FT /function="DNA POLYMERASE III IS A COMPLEX, MULTICHAIN
 FT ENZYME RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN
 FT BACTERIA. THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5'

FT EXONUCLEASE ACTIVITY. THE BETA CHAIN IS REQUIRED FOR
 FT INITIATION OF REPLICATION ONCE IT IS CLAMPED ONTO DNA, IT
 FT SLIDES FREELY (BIDIRECTIONAL AND ATP-INDEPENDENT) ALONG
 FT DUPLEX DNA [CATALYTIC ACTIVITY: N deoxynucleoside
 FT triphosphate = N diphosphate + {DNA}[n]."
 FT /gene="dnan"
 FT /locus_tag="RV0002"
 FT /product="DNA POLYMERASE III (BETA CHAIN) DnaN (DNA
 FT NUCLEOTIDYLTRANSFERASE)"
 FT /translation="MDAATRVGLTDTFRLRLRESFADAVSWAKNLPARPAPVPLVSGV
 FT LITCGNGLITSGFDYVSAEQVAGVLSVNSGRLLSDITRALPNKPVDDVHVVG
 FT NVALTTCGNARPSLPVPEVDYPLPTLPETGLPAELFAELSAIQVAIGRDDTLPM
 FT LTGRVILGETVVAADTRFRAIRLKKSSASSPDEAAVLVPKILAAAKAGGGS
 FT DVRLSLGTGPGVKGDLGSGNGKSTTLLDRAFFPKFQLLPTEHTATVMDVAELI
 FT EAIKLAVADRGAQVRFADSGVLSAGADVDVRAEEDLVVDYAGEPTIAFNFTYL
 FT TDGLSSLSRVSFSGFTTAGKPAALLRPVSGDRPVAGLNGGPFPAVSTDYVLLMPVR
 FT LPG"
 FT 3280..4437
 FT CDS
 FT /evidence=EXPERIMENTAL
 FT /note="Rv0003, (MTCY10H4.01), len: 385 aa. recP, DNA
 FT replication and repair protein (see citations below),
 FT equivalent to others Mycobacterial DNA replication and
 FT repair proteins e.g. NP 301131.1|NC 002677 from
 FT Mycobacterium leprae (385 aa); Q9L7L5|RECF_MYCPA from
 FT Mycobacterium avium subsp. paratuberculosis (385 aa);
 FT P50916|RECF_MYCSM from Mycobacterium smegmatis (384 aa);
 FT etc. Also highly similar to others e.g. P36176|RECF_STRCO
 FT DNA REPLICATION AND REPAIR PROTEIN from Streptomyces
 FT coelicolor (373 aa); NP 440892.1|NC 000911 from
 FT Synecocystis sp. strain PCC 6803 (384 aa);
 FT NP 469352.1|NC 003212 from Listeria innocua (370 aa); etc.
 FT Contains P50017 ATP/GTP-binding site motif A (P-loop).
 FT P50617 RecF protein signature 1, and P500618 RecF protein
 FT signature 2. BELONGS TO THE RECF FAMILY."
 FT /transl_table=11
 FT /function="THE RECF PROTEIN IS INVOLVED IN DNA METABOLISM
 FT AND RECOMBINATION; IT IS REQUIRED FOR DNA REPLICATION AND
 FT NORMAL SOS INDUCIBILITY. RECF BINDS PREFERENTIALLY TO
 FT SINGLE-STRANDED, LINEAR DNA. IT ALSO SEEMS TO BIND ATP."
 FT /gene="recF"
 FT /locus_tag="RV0003"
 FT /product="DNA REPLICATION AND REPAIR PROTEIN RECF
 FT (SINGLE-STRAND DNA BINDING PROTEIN)"
 FT /translation="MYVRLHGLDRFSWACVLELHPGRTVFVPGNGYKTNLEALWY
 FT STTLGSHRSVADLPIRVGTDRAVISTIVNDGREGCAVDLEIATGRVKNLARNSVRS
 FT TRDVGVLRVILFAPDLGLVGRDPADRRYLDLAIVRPAIAAARAEVERVLRQTA
 FT LKXSPGARVGRGVFTLLEVDSDSLAEHGAELVAARDLVNQLAPEVKKAKQLLAPE
 FT SRASIGYRASMVDVTFSEQSDIRQLAAALRLAARDAELERGVCLVGGPHRDDLI
 FT LRLGDQPAKGFGASHGEASLVALRLAAYQLLRVDGEPVLLDQVDFEALDVMRRRLA
 FT TAAESARQVLVTAALVLEIDIPAGDARVRHIDVRADDTGSMVVLP"
 FT 3367..3390
 FT /note="P500017 ATP/GTP-binding site motif A"
 FT 3634..3690
 FT /note="P500617 RecF protein signature 1"
 FT 4243..4296
 FT /note="P500618 RecF protein signature 2"
 FT 4434..4997
 FT CDS
 FT /evidence=EXPERIMENTAL
 FT /note="Rv0004, (MTCY10H4.02), len: 187 aa. Conserved
 FT hypothetical protein (see Salazar et al., 1996), highly
 FT similar, but longer 21 aa in N-terminus, to
 FT AAV33636.1|AF222789 unknown protein from Mycobacterium
 FT avium subsp. paratuberculosis (166 aa); and highly similar
 FT to NP 301132.1|NC 002677 conserved hypothetical protein
 FT from Mycobacterium leprae (189 aa); S70990 hypothetical
 FT protein from Mycobacterium smegmatis (194 aa). Also highly
 FT similar, except in N-terminal part, to

Query Match 100.0%; Score 24; DB 15; Length 341957;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACCGTGGCTGGATGACGGTCTC 24
 Db 55594 GACCGTGGCTGGATGACGGTCTC 55571

RESULT 4
 BX248334/c
 LOCUS
 DEFINITION BX248334 343050 bp DNA linear BCT 11-JUN-2003
 Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment 1/14.

ACCESSION BX248334 BX248333
 VERSION BX248334.1 GI:31616762
 KEYWORDS complete genome.
 SOURCE Mycobacterium bovis subsp. bovis AF2122/97
 ORGANISM Mycobacterium bovis subsp. bovis AF2122/97
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
 tuberculosis complex.

REFERENCE 1
 AUTHORS Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H.,
 Pryor, M., Duttoy, S., Grondin, S., Lacroix, C., Monsemp, C., Simon, S.,
 Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L.,
 Wheeler, P. R., Parkhill, J., Barrall, B. G., Cole, S. T., Gordon, S. V. and
 Hewinson, G.
 The complete genome sequence of Mycobacterium bovis
 Online Publication
 PNAS 10.1073/pnas.1130426100 (Microbiology)
 2 (bases 1 to 343050)

TITLE Direct Submission
 JOURNAL Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
 Bacterienne Institut Pasteur 28, rue du Dr Roux 75724 PARIS cedex
 15, France. e-mail: tgarnier@pasteur.fr Submitted on behalf of the
 Mycobacterium bovis sequencing teams. TB Research Group, Veterinary
 Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,
 Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA, UK. FTA Annotation, Genopole, Institut
 Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
 Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
 rue du Docteur Roux, 75724 Paris Cedex 15, France

FEATURES
 source
 1. 343050
 /organism="Mycobacterium bovis subsp. bovis AF2122/97"
 /mol_type="genomic DNA"
 /strain="AF2122/97"
 /db_xref="taxon:233413"
 1. 1524
 /gene="dnaA"
 /locus_tag="Mb00001"
 1. 1524
 /gene="dnaA"
 /locus_tag="Mb00001"
 /notes="Mb00001", dnaA, len: 507 aa. Equivalent to RV0001,
 len: 507 aa, from Mycobacterium tuberculosis strain H37Rv,
 (99.6% identity in 507 aa overlap). dnaA, chromosomal
 replication initiator protein (see citations below),
 equivalent to other Mycobacterial CHROMOSOMAL REPLICATION
 INITIATOR PROTEINS e.g. P46388|DnaA MYCLE from
 Mycobacterium leprae (502 aa); Q9L7L7|DnaA MYCPA from
 Mycobacterium paratuberculosis (509 aa); P49990|DnaA MYCAV
 from Mycobacterium avium (508 aa); P49992|DnaA MYCSM from
 Mycobacterium smegmatis (504 aa); etc. Also highly similar
 to others except in N-terminus e.g. Q9ZH75|DnaA STICH
 CHROMOSOMAL REPLICATION INITIATOR PROTEIN from
 Streptomyces chrysomallus (524 aa); Q9ZH76|DnaA STREE from
 Streptomyces reticuli (643 aa); DnaA ECOLI|P03004|B3702
 chromosomal replication initiator protein from Escherichia
 coli strain K12 (467 aa). FASTA scores: opt: 986. E(1): 0,
 (43.2% identity in 389 aa overlap); etc. Contains P50017
 ATP/GTP-binding site motif A (P-loop) and P50108 DnaA
 protein signature. BELONGS TO THE DnaA FAMILY. Note that
 the first base of this gene has been taken as base 1 of

the Mycobacterium bovis genomic sequence."

/codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="CHROMOSOMAL REPLICATION INITIATOR PROTEIN DnaA"
 /protein_id="CAD92863.1"
 /db_xref="GI:31616763"
 /translation="MTDPSGFTTVMNVAVSELNGPKYDDGSSDANLSAPLTPOQ
 PAKNLIQVPIVEGFEALLSVBSFVQNEIERHLRAPITDALSPRLGHQIOLGVRIAP
 PATADADITVPPSENATPSTDTTNDSDIDSAARGNQHSWPFYFTEPRNTDS
 ATAGTSLNRVITDTEFVIGASNRFAHAALAAEAARAYNPLFIWGESGLKTHLL
 HAAGNYAQRLLPFGNRVSTEFNTDINSRDKRVAFKRSYRDVLLVDIQFI
 EGKEGIQEFPHFTNTLNANKQIVISDRPKQLATLELRTRFVLLVDVQPE
 LETRIALRKQAMERLAI PDDVLELIASSIRNIREGALIRVTFAPASLNTKPIDK
 ALAEIARDLADANTMQISAATIMATAEYEDTVEELRPGKTRALASQROFAMVL
 CRELTDLSLKPIGQAFGRDHTVMYAKRKILSEAEERREYDFHVKELTTRIQRSKR"
 2052..3260
 /gene="dnaA"
 /locus_tag="Mb00002"
 2052..3260
 /gene="dnaA"
 /locus_tag="Mb00002"
 /EC_number="2.7.7.7"
 /note="Mb00002, dnaN, len: 402 aa. Equivalent to RV0002,
 len: 402 aa, from Mycobacterium tuberculosis strain H37Rv,
 (99.8% identity in 402 aa overlap). dnaN, DNA polymerase
 III (beta chain) (EC 2.7.7.7) (see citations below),
 equivalent to other Mycobacterial DNA POLYMERASES III BETA
 CHAIN e.g. NP_301130.1|NC_002677 from Mycobacterium leprae
 (399 aa); Q9L7L6|DP3B MYCPA from Mycobacterium avium
 subsp. paratuberculosis (399 aa); P52851|DP3B MYCSM from
 Mycobacterium smegmatis (397 aa); etc. Also highly similar
 to others e.g. P27903|DP3B STROCO DNA POLYMERASE III BETA
 CHAIN from Streptomyces coelicolor (376 aa), FASTA scores:
 opt: 1189. E(1): 0, (52.8% identity in 337 aa overlap);
 P21174|DP3B MICL from Micrococcus luteus (310 aa);
 P52223|DP3B SYN7 from Synecococcus sp. strain PCC 7942
 (375 aa); etc. Overlaps and extends CDS in neighbouring
 cosmid MTCY10H4.01."
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="DNA POLYMERASE III (BETA CHAIN) DnaN (DNA
 NUCLEOTIDYLTRANSFERASE)"
 /protein_id="CAD92864.1"
 /db_xref="GI:31616764"
 /translation="MDAATRVGLTDLTRLLRESFADAVSWAKNLPAPFVPLSG
 VLTGSDNGLTISGFDYEAQVAGAEIVSGSVLVSGRLSDITRALPNKPVGVHV
 BGNRVALTGNARFSLPTMVEDYPTLPTLPEETGLPAELFAEIAISOVALAAGRDVT
 LPMLTGIRVEILGHTVLAVDREPLAVRELKWSASSPDLEAVLPAKTLAEAKAG
 IGGSDVRISLGTGFGVKGKGLLGSIGNKSTTRLLDAEFKFLQLPTHTATVATMD
 VAELEIAKLVALVADRAQVMEFADGVSLGADDDVGRAEEDLVVDYAGEPITIA
 FNTYLTDLGLSSLSERSVFGFTTAGKPAALLRPVSGDDRPVAGLNGNGPFPVSTDYV
 YLLMPVRLPG"
 3280..4437
 /gene="recF"
 /locus_tag="Mb00003"
 3280..4437
 /gene="recF"
 /locus_tag="Mb00003"
 /locus_tag="Mb00003"
 /note="Mb00003, recF, len: 385 aa. Equivalent to RV0003,
 len: 385 aa, from Mycobacterium tuberculosis strain H37Rv,
 (99.5% identity in 385 aa overlap). recF, DNA replication
 and repair protein (see citations below), equivalent to
 others Mycobacterial DNA replication and repair proteins
 e.g. NP_301131.1|NC_002677 from Mycobacterium leprae (385
 aa); Q9L7L5|RECF MYCPA from Mycobacterium avium subsp.
 paratuberculosis (385 aa); P50916|RECF MYCSM from
 Mycobacterium smegmatis (384 aa); etc. Also highly similar
 to others e.g. P36176|RECF STROCO DNA REPLICATION AND
 REPAIR PROTEIN from Streptomyces coelicolor (373 aa);
 NP_440892.1|NC_000911 from Synecococcus sp. strain PCC
 6803 (384 aa); NP_469352.1|NC_003212 from Listeria innocua

Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J., Mizuno,K., Nariawa,S., Nishikawa,J., Okada,M., Ryu,R., Sugano,S., Sugiyama,A., Suzuki,Y., Tsumoda,Y., Ueda,M., Xie,Q., Yokomizo,S., Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J., Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N., Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Sato,K., Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K., Yasunishi,A. and Hayashizaki,Y.

FEATURES

source
1..2537
Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="002-110-H08"

ORIGIN

Query Match 80.0%; Score 19.2; DB 8; Length 2537;
Best Local Similarity 87.5%; Pred. No. 2.4e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GACCGTGGCTGGATGACGGTCTC 24
|||||
DB 1758 GACCGTGGCTGGATGCGGCACTC 1781

RESULT 6

AP005251/c
LOCUS 67000 bp DNA linear PLN 27-FEB-2003
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8,
BAC clone:OSJNB0011H15, complete sequence.

ACCESSION AP005251
VERSION AP005251.2 GI:28569994
KEYWORDS HTG.

SOURCE

Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

1 Sasaki,T., Matsumoto,T. and Katayose,Y.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC
clone:OSJNB0011H15
Published Only in Database (2002)

JOURNAL

REFERENCE 2 (bases 1 to 67000)
AUTHORS Sasaki,T., Matsumoto,T. and Katayose,Y.
JOURNAL Submitted (23-MAY-2002) Takuji Sasaki, National Institute of
Agricultural Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)
On Feb 26, 2003 this sequence version replaced gi:21165567.
The orientation of the sequence is from M13rev to -21M13 of the BAC
clone.

FEATURES

source
1..67000
Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="8"
/clone="OSJNB0011H15"

ORIGIN

Query Match 80.0%; Score 19.2; DB 8; Length 67000;
Best Local Similarity 87.5%; Pred. No. 9.3e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GACCGTGGCTGGATGACGGTCTC 24
|||||
DB 39877 GACCGTGGCTGGATGGCGCACTC 39854

RESULT 7

AP003893
LOCUS 71195 bp DNA linear HTG 21-MAR-2002
DEFINITION Oryza sativa (japonica cultivar-group) chromosome 8 clone
CU1364_H06, *** SEQUENCING IN PROGRESS ***.

ACCESSION AP003893.1 GI:14646851
VERSION AP003893
KEYWORDS HTG; HTGS PHASE2.

SOURCE

Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

1 Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, BAC
clone:CU1364_H06

JOURNAL

REFERENCE 2 (bases 1 to 71195)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
JOURNAL Direct Submission
Submitted (09-JUL-2001) Takuji Sasaki, National Institute of
Agricultural Sciences, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT

The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

source
1..71195
Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="8"
/clone="OU1364_H06"

ORIGIN

Query Match 80.0%; Score 19.2; DB 2; Length 71195;
Best Local Similarity 87.5%; Pred. No. 9.2e+02;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GACCGTGGCTGGATGACGGTCTC 24
|||||
DB 5479 GACCGTGGCTGGATGGCGCACTC 5502

RESULT 8

SCWHIE
LOCUS 5749 bp DNA linear BCT 15-APR-1991
DEFINITION S. coelicolor white locus DNA.
ACCESSION X55942

```

VERSION X55942.1 GI:46919
KEYWORDS whole locus.
SOURCE Streptomyces coelicolor
ORGANISM Streptomyces coelicolor
          Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
          Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 5749)
AUTHORS Davis,N.K. and Chater,K.F.
TITLE Spore colour in Streptomyces coelicolor A3(2) involves the
        developmentally regulated synthesis of a compound biosynthetically
        related to polyketide antibiotics
JOURNAL Mol. Microbiol. 4 (10), 1679-1691 (1990)
MEDLINE 91171868
PUBMED 2077356
REFERENCE 2 (bases 1 to 5749)
AUTHORS Bruton,C.J., Chater,K.F. and Davis,N.K.
TITLE Direct Submission
JOURNAL Submitted (18-SEP-1990) C.J. Bruton, K.F. Chater and N.K. Davis,
        JOHN INNES INSTITUTE, JOHN INNES CENTRE FOR PLANT SCIENCE RESEARCH,
        COLNEY LANE, NORWICH NR4 7UH, UK
FEATURES
     source
       1..5749
         Location/Qualifiers
           /organism="Streptomyces coelicolor"
           /mol_type="genomic DNA"
           /db_xref="taxon:1902"
           167..177
             RBS
             gene
               185..1378
                 /gene="ORF I"
                 <185..>1378
                   mRNA
                   CDS
                     185..1378
                       /gene="ORF I"
                       /codon_start=1
                       /transl_table=11
                       /protein_id="CAA39406.1"
                       /db_xref="GI:46920"
                       /db_xref="SWISS-PROT:P23158"
                       /translations="MTVSPVATDAPSTDTATTSATSPAVTAGGVTSAPFGS
RVEVLMVDHGMQEFIDAVIRDRVAAPVGHVSDCOLCSLENPTOMLTSSES
AAPFLAWNSDEHLDVEPLACVDRTHSLRYSVLRETGGPAPGERSAPRIGDNV
VRALFTVRPGTEASTRLSEYVSPDAHVGSTRLRLSLFNSGNRIIVRAVEVGD
LQALRHVAPQGVRAVEEALNPYLEQDLGDPOSARFFTRAAPVPHATYDPS
GAREERLALYPVRDAGDPLARLLAQDAAARNPDGPVLAATVFRDLDLVRLVDV
DGPEDAPAEVLGLHGGAADAERLLDAAAAGVDGSPAEAAATLSRLRLRTRTLDTR
RSGS"
                     mat_peptide
                       185..1375
                         /gene="ORF I"
                         /product="unnamed"
                         <185..>1378
                           prim_transcript
                           misc_feature
                             1383..1439
                               /note="GC rich region"
                               /note="strand-specific purine-pyrimidine bias region"
                               misc_feature
                                 1450..1504
                                   RBS
                                   gene
                                     1513..1519
                                       /gene="ORF II"
                                       <1525..>1598
                                         mRNA
                                         CDS
                                           1525..1598
                                             /gene="ORF II"
                                             /codon_start=1
                                             /transl_table=11
                                             /protein_id="CAA39407.1"
                                             /db_xref="GI:46921"
                                             /db_xref="SWISS-PROT:P23157"
                                             /translations="MTDQQRIVLSGLDIAPNRERGDLRALITPTAGTSQFMGVAI
VPGDRIAEHYHPSEFYVTEGALAEVLDVGPHSLGTGQGLLIPQMRHRFRNAGD
VEARLVHGLGAPRPELGHVDTEADRGNALPCDGAHAGHAGHERPAPAEVVS"
                                           mat_peptide
                                             1525..1595
                                               /gene="ORF II"
                                               /product="unnamed"
                                               <1525..>1598
                                                 prim_transcript

```

```

RBS
gene
mRNA
CDS
/ gene="ORF II"
1984..1989
/ gene="ORF II"
1995..3266
/ gene="ORF III"
<1995..>3266
/ gene="ORF III"
1995..3266
/ gene="ORF III"
/ gene="ORF III"
/ codon_start=1
/ transl_table=11
/ product="putative B-ketoacyl synthase"
/ protein_id="CAA39408.1"
/ db_xref="GI:46922"
/ db_xref="GOA:P23155"
/ db_xref="SWISS-PROT:P23155"
/ translations="MTRRRVAVTGIVGVPFGIGTPOFWLLSGEATRRISLFDPS
GLRQIAECDFEPSDHDGLGLATACRCDRVVQCPALVAASAVRDANLDMREDPWRAG
ATLTAVTGGTIRLEHDYVLYSERGSRWDVDRSEPHLEAFPTATLSSVAEFGVR
GPVTVSGTISGLDVAIYAHVAGRGVDVCLAGAADSPISITMACFDATKATSN
NDPFAHASRPFDADNRGFMVMEGAALVLDLEHARARGADVYCEVGYATFGNAVHM
TGLTKEGLEMARAIIDTALDMLDGLSDIDVNAHSGTQQRDRHETAAVRSRSHAY
ATPMSSIKSMVGHSLGAIGSTIELAACVLAHAHVPTANTYTPDPCDLDLYVPREAR
ERTLRHVLVSGSGFGFQSAVLSGSGGLR"
1995..3263
/ gene="ORF III"
/ product="putative B-ketoacyl synthase"
<1995..>3266
/ gene="ORF III"
3250..3256
/ gene="ORF III"
3263..4537
/ gene="ORF IV"
<3263..>4537
/ gene="ORF IV"
3263..4537
/ gene="ORF IV"
/ codon_start=1
/ transl_table=11
/ protein_id="CAA39409.1"
/ db_xref="GI:46923"
/ db_xref="GOA:P23156"
/ db_xref="SWISS-PROT:P23156"
/ translations="MSGPQRTGTGGSSRAVVTGLVLSPHGTGVEAHKAVADGTS
LGPVTRGGCAHLPLRVAGVGHGDAETVEDRFLVOTDRFTHALSATOHALADARFG
RADVDSPYGVVVTAGSGGEGFQRELQNLWGHSGRHVGPYOSIAFWAATGQVSI
RNDKPGCGVVADEAGGLDALAHALAVNGTDTVVCGATEAPLAPYSIVCQIGYPE
LSRATEPDRAIRPFTAAACGFAFGAGVAVVBEAAARBERGADVAPVAGHAATFG
AGRWAESEGLARAIQGAALAEACGRPEEDVVFDALGVPEADRALALADALGPHA
ARVETAPKTGTGRAYCAAPVLDVATVATLAMEHGLIIPPTPHVLDVCHDLDTVGRAP
AEPTALVLARGLMGSNSALVLRGAVPPEGR"
3263..4534
/ gene="ORF IV"
/ product="unnamed"
<3263..>4537
/ gene="ORF IV"
4538..4605
/ note="intergenic region"
4593..4599
4606..4878
/ gene="ORF V"
<4606..>4878
/ gene="ORF V"
4606..4878
/ gene="ORF V"
/ codon_start=1
/ transl_table=11
/ product="putative acyl carrier"
/ protein_id="CAA39410.1"
/ db_xref="GI:46924"
/ db_xref="SWISS-PROT:P23153"
/ translations="MTDQQLDQVTVVEELSALMKRTAGVHVDPTVLRQQADDGDFTFG
LDSLGLIGIVAELEKRYGLPQAEBCRCKTPADFLALVNGALKTV"

```

```

mat_peptide      4506. .4875
                  /gene="ORF V"
                  /product="putative acyl carrier"
prim_transcript  <4506. .>4878
                  /gene="ORF V"
RBS              4865. .4874
                  /gene="ORF V"
gene             4880. .5359
                  /gene="ORF VI"
mRNA             <4880. .>5359
                  /gene="ORF VI"
CDS              4880. .5359
                  /gene="ORF VI"
                  /codon_start=1
                  /transl_table=11
                  /product="putative polyketide cyclase"
                  /protein_id="CAA39411.1"
                  /db_xref="GI:46925"
                  /db_xref="GOA:P23154"
                  /db_xref="SWISS-PROT:P23154"
                  /translation="MAGHTDNRITTAAPMELVNMVNDIEKPGLFSEYASVEVLGRD
DKVYFRLTHPDADGKYSWVSERVADPVRIVRAQRVETGTFQYQNMIIWYASVTAET
GIVMWTDQFAMKDPADVDAMWTDNINRNSRTQMALINDRIEQAGERTASVLAD"
                  >5356
mat_peptide      /gene="ORF VI"
                  /product="putative polyketide cyclase"
prim_transcript  <4880. .>5359
                  /gene="ORF VI"
misc_feature     5360. .5392
                  /note="intergenic region"
RBS              5378. .5385
gene             5393. .5728
                  /gene="ORF VII"
mRNA             <5393. .>5728
                  /gene="ORF VII"
CDS              5393. .5728
                  /gene="ORF VII"
                  /codon_start=1
                  /transl_table=11
                  /protein_id="CAA39412.1"
                  /db_xref="GI:46926"
                  /db_xref="GOA:P23159"
                  /db_xref="SWISS-PROT:P23159"
                  /translation="MHHTLIVARMAFGAAPDIKAVFASDSGELPHLVGNRRSLFEF
GCGVYHLIESDEDPATTIGRLTGHPEFRQVSERLEPVVSAYDPATWRGPKDAMARCF
YRWERTPAG"
mat_peptide      5393. .5725
                  /gene="ORF VII"
                  /product="unnamed"
prim_transcript  <5393. .>5728
                  /gene="ORF VII"

Query Match      76.7%; Score 18.4; DB 1; Length 5749;
Best Local Similarity 95.0%; Pred. No. 3.9e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 5 GTGGCGTGATGACGGTCTC 24
    ||||| ||||| ||||| |||||
Db 176 GTGGCGCGATGACGGTCTC 195

RESULT 9
SC039123/c      300100 bp      DNA      linear      BCT 11-FEB-2003
LOCUS           Streptomyces coelicolor A3(2) complete genome; segment 20/29.
DEFINITION      Streptomyces coelicolor A3(2) complete genome; segment 20/29.
ACCESSION       AL031184 AL035569 AL079356 AL137242 AL138978 AL157916
AL359152 AL359214 AL391588 AL391763 AL589707 AL645882
VERSION         AL939123.1 GI:24430032
KEYWORDS
SOURCE          Streptomyces coelicolor A3(2)
ORGANISM        Streptomyces coelicolor A3(2)
                Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
                Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
AUTHORS         Bentley,S.D., Chater,K.F., Cerdeno-Tarraga,A.M., Challis,G.L.,
                Thomson,N.R., James,K.D., Harris,D.E., Quail,M.A., Kieser,H.,
                Harper,D., Bateman,A., Brown,S., Chandra,G., Chen,C.W., Collins,M.,
                Cronin,A., Fraser,A., Goble,A., Hidalgo,J., Hornsby,T., Howarth,S.,
                Huang,C.H., Kieser,T., Lake,L., MurphY,L., Oliver,K., O'Neill,S.,
                Rabinowitsch,E., Rajandream,M.A., Rutherford,K., Rutter,S.,
                Seeger,K., Saunders,D., Sharp,S., Squares,R., Squares,S.,
                Taylor,K., Warren,T., Wietzorrek,A., Woodward,J., Barrall,B.G.,
                Parkhill,J. and Hopwood,D.A.
                Complete genome sequence of the model actinomycete Streptomyces
                coelicolor A3(2)
                Nature 417 (6885), 141-147 (2002)
JOURNAL          21996410
MEDLINE          12000953
PUBMED
AUTHORS         Bentley,S.D.
TITLE           Direct Submission
JOURNAL
COMMENT          Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
                sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
                Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
                On or before Oct 30, 2002 this sequence version replaced
                gi:20520687, gi:20520786, gi:20520826, gi:20520797, gi:20520798,
                gi:20520697, gi:20520859, gi:20520743, gi:20520862.
                Location/Qualifiers
                1. 300100
                /organism="Streptomyces coelicolor A3(2)"
                /mol_type="genomic DNA"
                /strain="A3(2)"
                /db_xref="taxon:100226"
                53. 56
                66. 1151
                /gene="SC05222"
                /note="synonym: SC7E4.19"
                66. 1151
                /gene="SC05222"
                /note="SC05222"
                /note="SC7E4.19, possible lyase, len: 361 aa, similar to
                SW:PTLS_S7RSQ (EMBL:U05213) Streptomyces sp. pentalenene
                synthase (EC 4.6.1.5), 336 aa; fasta scores: Opt: 254
                z-score: 317.5 E(): 3.3e-10; 23.8% identity in 311 aa
                overlap. Contains match to Prosite entry PS00904 Protein
                prenyltransferases alpha subunit repeat signature and a
                TTA leucine codon, possible target for bldA regulation"
                /codon_start=1
                /transl_table=11
                /product="putative lyase"
                /protein_id="CA394607.1"
                /db_xref="GI:8546888"
                /db_xref="GOA:Q9K499"
                /db_xref="SPTREMBL:Q9K499"
                /translation="MHAFPHGTATPTAIVPPSLRPLVIEAAPPRLQHPYWPKLQET
                TRTWLEKRLMPADKVEYADGLCYTDMAGYLGAFDEVLQAIADYASNFVWDDRH
                DRDIVHGRAGARRLRGLLTHALDPSGDHLHEDTLVAGFADSVRLYAFLPATWNR
                FARPHVTIEAYDRFENRTGRIVGVVEYLELRLLTAFAHWITDLELPPSGCGLPDA
                VRKHPAYRRALLSCEFAAWNDLCSLPKETAGDEVNGLSLTHSLTLEEALGEV
                RRVRECIITTEFLAVDELADGTGKLSGAVRANVGNMRWFSVYVWFH
                BSRGIMVDSWDDRSITPFFVNNAAAGEK"
                327. 356
                /gene="SC05222"
                /note="PS00904 Protein prenyltransferases alpha subunit
                repeat signature"
                948. .950
                /gene="none"
                948. .950
                /gene="none"
                /note="TTA leucine codon. Possible target for bldA
                regulation"
                1148. 2533
                /gene="SC05223"
                /note="synonym: SC7E4.20"
                1148. 2533
                /gene="SC05223"
                /note="SC05223"
                /note="SC05223"
                /note="SC7E4.20, probable cytochrome P450, len: 461 aa;

```

similar to SW:CP51_HUMAN (EMBL:U23942) Homo sapiens
 cytochrome P450 51-EC 1.14.14.1) Cyp51, 503 aa; fasta
 scores: opt: 532 z-score: 587.2 E(): 3.1e-25; 25.7%
 identity in 447 aa overlap. Contains Pfam match to entry
 PF00067 p450, Cytochrome P450"

/codon_start=1
/transl_table=11
/product="putative cytochrome P450"
/protein_id="CAB94608.1"
/db_xref="GI:8546889"
/db_xref="GOA:Q9K498"
/db_xref="SPTRMBL:Q9K498"
/translation="MTVESVNPETRAPAPAGAPDELPPVAGGVLLGHGWRLLARDP
LAFMSQLRDGVDVRIKIGPKTYAVTNPELTGALALNPDIYHAGPQWLESLEGLGKE
GVATANGLRQRRITQAPRLDAI PAYGIMEEHALTERQPGKTVDTASFSFR
VAVRVAARCLLRQYMDERAERLCVALATVFRGMYRMVPLGPLYRLPLPANRRFND
ALADLLVDEIIAERRASQKPDLLTALLEAKDNDGDP IGEQETHDOVVAITPFGS
ETASITMQLQALADHPHADRIDEVAATGGRVAPEDVKEKLRHTGNVIVEAMRL
RPAVWLTTRAVAESLGLYRIAGADIISYPYAIQRDKSYDDNLEFPDPDWLPERA
ANPKYAMKFFSAGKRCPSDFHFSMAQLITLITAAALATKYRFQVAGSDAVRVGITLR
PHDLLVRPVAR"
1217..2428
/genes="SC05223"
/notes="Pfam match to entry PF00067 p450, Cytochrome P450,
score 221.80, E-value 1e-62"
/complement(2547..3515)
/genes="SC05224"
/notes="synonym: SC7E4.21c"
/complement(2547..3515)
/genes="SC05224"
/notes="SC05224"
/notes="SC7E4.21c; probable araC-family transcriptional
regulator, len: 322 aa; similar to TR:Q9S166
(EMBL:AB023785) Streptomyces griseus transcriptional
activator for strR, AdpA, 405 aa; fasta scores: opt: 1005
z-score: 1177.5 E(): 0; 49.2% identity in 325 aa
overlap and to TR:CAB87229 (EMBL:AL163641) Streptomyces
coelicolor araC-family transcriptional regulator AdpA, 398
aa; fasta scores: opt: 990 z-score: 1160.1 E(): 0; 48.9%
identity in 313 aa overlap. Contains Pfam match to entry
PF00165 HTH_Arac, Bacterial regulatory helix-turn-helix
proteins, araC family and match to Prosite entry PS00041
Bacterial regulatory proteins, araC family signature.
Contains also a possible helix-turn-helix motif at
residues 230..251 (+3.72 SD)"
/codon_start=1
/transl_table=11
/product="putative araC-family transcriptional regulator"
/protein_id="CAB94609.1"
/db_xref="GI:8546890"
/db_xref="GOA:Q9K497"
/db_xref="SPTRMBL:Q9K497"
/translation="MLRNVAALLDGAHPFELGVCEVFGIDRSDEGLPVYFVAVSA
EGPLTHVGLVSTPYGLERLEADLIAPAGSDFVEREYPPDLLALREARTDGT
RVLSVGVFVLGAGLLGRCQCAVHWAELARQHPARVAPVLYVDEPVTISA
GTAAGIDACLVHVRKEQSEVARIARVPPHRRDGGQQAQIERPLPSSCDTVGEV
LAWMEQHLDEVEQGLAVRAHSPRTFARRFQQTGTPPYRMILQRVYLLAQRLLA
TDTMTDTAWRAGFGTAALRHQFTALDITTHAVRTRPGPEAAA"
complement(2571..2831)
/genes="SC05224"
/notes="Pfam match to entry PF00165 HTH_Arac, Bacterial
regulatory helix-turn-helix proteins, araC family, score
79.70, E-value 6e-20"
/complement(2592..2720)
/genes="SC05224"
/notes="PS00041 Bacterial regulatory proteins, araC family
signature"
/complement(3557..7543)
/notes="previously sequenced DNA fragment. EMBL:AJ276618
Streptomyces coelicolor A3(2) nrdX gene, nrdL gene and
nrdM gene"
/complement(3787..4818)
/genes="SC05225"
/notes="synonyms: nrdM, SC7E4.22c"

CDS
complement(3787..4818)
/genes="SC05225"
/notes="SC7E4.22c, nrdM, ribonucleotide-diphosphate
reductase small chain, len: 343 aa; identical to
previously sequenced TR:CAB82486 (EMBL:AJ276618)
Streptomyces coelicolor A3(2) ribonucleotide-diphosphate
reductase small chain (EC 1.17.4.1) nrdM, 343 aa. Contains
Pfam match to entry PF00268 ribonuc red, Ribonucleotide
reductases and a possible hydrophobic membrane spanning
region"
/codon_start=1
/transl_table=11
/product="ribonucleotide-diphosphate reductase small
chain"
/protein_id="CAB94610.1"
/db_xref="GI:8546891"
/db_xref="GOA:Q9K3C2"
/db_xref="SPTRMBL:Q9K3C2"
/translation="MTTETPAHTTKNLLDPGELTLRPMRYPDFVEYVDAIKNTWH
VEVDLSADVADLAKLSFMEQHLIGRLVAFAGDSIVANNLILIKHINSPEARLY
LSRQLFEAAHVQFYLLDITPDPEADRAAFAAVENPISREKAEFCFRWSDVES
IDRLKETADRRRLNLICFAACIEGLFFYCAFYVYFRSGLGLHGLATGTNWVFD
ETMWSFAFDVDTVRKEPELFDQLREQVTMLREAVAEALQFADLQCDGLPGMN
TDSMRYLECVADQRLTRLGFPVYGSFPMELQGVQELTNFERRPSAYAVE
GTVDLSEDF"
complement(3832..4592)
/genes="SC05225"
/notes="Pfam match to entry PF00268 ribonuc red,
Ribonucleotide reductases, score 222.30, E-value 1e-75"
/complement(4818..7193)
/genes="SC05226"
/notes="synonyms: nrdL, SC7E4.23c"
/complement(4818..7193)
/genes="SC05226"
/notes="SC7E4.23c, nrdL, ribonucleotide-diphosphate
reductase large chain, len: 791 aa; identical to
previously sequenced TR:CAB82485 (EMBL:AJ276618)
Streptomyces coelicolor A3(2) ribonucleotide-diphosphate
reductase large chain (EC 1.17.4.1) nrdL, 791 aa. Contains
Pfam match to entry PF00317 ribonucleo red, Ribonucleotide
reductase and match to Prosite entry PS00089
Ribonucleotide reductase large subunit signature"
/codon_start=1
/transl_table=11
Query Match 76.7%; Score 18.4; DB 1; Length 300100;
Best Local Similarity 95.0%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 5 GTGGGCTGGATGACGGTCTC 24
|||||
Db 110356 GTGGGCGGATGACGGTCTC 110337
|||||
RESULT 10
AY335636/c
LOCUS
DEFINITION
AY335636
939 bp mRNA linear SYN 15-OCT-2003
Synthetic construct Homo sapiens pyridoxal kinase (PDXK) mRNA,
partial cds.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AY335636.1 GI:33303898
FLI_CDNA.
synthetic construct
synthetic construct
artificial sequences.
1 (bases 1 to 939)
Park, J., Rolfs, A., Hu, Y., Shen, B., Vannberg, F., Moreira, D.,
Kelley, T., Zuo, D., Raphael, J., Baqui, M., Jepson, D., Harlow, E.,
Labae, J. and Brizuela, L.
Cloning of human full-length CDS FLEXGene kinases in
recombinational vector system
Unpublished
2 (bases 1 to 939)

AUTHORS Park, J., Rolfs, A., Hu, Y., Shen, B., Vannberg, F., Moreira, D., Kelley, T., Zuo, D., Raphael, J., Baqui, M., Jepson, D., Harlow, E., LaBaer, J. and Brizuela, L.

TITLE Direct Submission

JOURNAL Submitted (02-JUL-2003) Biological Chemistry and Molecular Pharmacology, Harvard Institute of Proteomics, 320 Charles St., Cambridge, MA 02141-2023, USA

COMMENT This CDS clone is a part of a collection of human full-length expression clones generated by Harvard Institute of Proteomics. Each CDS has been cloned without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion (TM) cloning system between the SalI and HindIII sites of the pDNR-Dual vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame.

FEATURES Location/Qualifiers

source 1..939

/organism="synthetic construct"

/mol_type="mRNA"

/db_xref="taxon:32630"

/clone="FLH000087.01L"

/clone_lib="WGC clone templates"

/lab_host="DH5alpha T1 resistant"

/notes="Vector: pDNR-Dual"

1..>939

/genes="PDXK"

1..>939

/genes="PDXK"

/notes="Mutations: Stop->Leu; pyridoxine, vitamin B6"

/codon_start=1

/transl_table=11

/product="pyridoxal kinase"

/protein_id="AAQ02463.1"

/db_xref="GI:33303899"

1..>939

/translation="MEEECRVLISIQSHVIRGVGNRAATFPQLVGLFEIDAVNSVQFS NHTGYAHKGVLSNDELQELYEGRLNNKDYVLTGYTRDKSFLAMVVDIVQELK QQNPLRVYCDPVLGDKWDEGSMVYEDLLPYKKEVPLADLITPNQFAELLSGR KHQSBEALRVMDLHMSGPDVTIVITSSLPQGSNYLIVLGSQRNRPAGSVVNER IRMDIRKVDVAVFGTGLFAAMLLAWTHKPNKLVACEKTVSTLHVLQRTIOCAKA QAGEGVRSPMQLELRMVQSKRDIEDPEIVVQATVLL"

gene

CDS

Query Match 75.8%; Score 18.2; DB 12; Length 939;
Best Local Similarity 87.0%; Pred. No. 7.7e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCGTGGCGTGGATGACGGTCTC 24
||||| ||||| ||||| ||||| |||||
Db 932 ACCGTGGCGTGGATGACGGTCTC 910

RESULT 11
AX410805/c
LOCUS
DEFINITION Sequence 3452 from Patent WO0229103.
ACCESSION AX410805
VERSION AX410805.1 GI:21443510
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1
Alvares, C., Horne, D., Peres-da-Silva, S. and Vockley, J.G.
Gene expression profiles in liver cancer
Patent: WO 0229103-A 3452 11-APR-2002;
GENE LOGIC INC (US)
Location/Qualifiers
1..960
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

FEATURES Location/Qualifiers

source 1..960

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

ORIGIN /note="EMBL/GenBank Accession No. U89606"

Query Match 75.8%; Score 18.2; DB 6; Length 960;
Best Local Similarity 87.0%; Pred. No. 7.7e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCGTGGCGTGGATGACGGTCTC 24
||||| ||||| ||||| ||||| |||||
Db 938 ACCGTGGCGTGGATGACGGTCTC 916

RESULT 12
HSU89606/c
LOCUS
DEFINITION Human pyridoxal kinase mRNA, complete cds.
ACCESSION U89606
VERSION U89606.1 GI:1946348
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 960)
Hanna, M.C., Turner, A.J. and Kirkness, E.F.
Human pyridoxal kinase. cDNA cloning, expression, and modulation by ligands of the benzodiazepine receptor
J. Biol. Chem. 272 (16), 10756-10760 (1997)
9099727
PUBMED
2 (bases 1 to 960)
Hanna, M.C., Turner, A.J. and Kirkness, E.F.
Direct Submission
Submitted (13-FEB-1997) Department of Molecular and Cellular Biology, The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA
Location/Qualifiers
1..960
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22"
7..945
/EC_number="2.7.1.35"
/function="required for synthesis of pyridoxal-5-phosphate from vitamin B6"
/codon_start=1
/product="pyridoxal kinase"
/protein_id="AAC51233.1"
/db_xref="GI:1946349"
/translation="MEEECRVLISIQSHVIRGVGNRAATFPQLVGLFEIDAVNSVQFS NHTGYAHKGVLSNDELQELYEGRLNNKDYVLTGYTRDKSFLAMVVDIVQELK QQNPLRVYCDPVLGDKWDEGSMVYEDLLPYKKEVPLADLITPNQFAELLSGR KHQSBEALRVMDLHMSGPDVTIVITSSLPQGSNYLIVLGSQRNRPAGSVVNER IRMDIRKVDVAVFGTGLFAAMLLAWTHKPNKLVACEKTVSTLHVLQRTIOCAKA QAGEGVRSPMQLELRMVQSKRDIEDPEIVVQATVLL"

ORIGIN

Query Match 75.8%; Score 18.2; DB 9; Length 960;
Best Local Similarity 87.0%; Pred. No. 7.7e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCGTGGCGTGGATGACGGTCTC 24
||||| ||||| ||||| ||||| |||||
Db 938 ACCGTGGCGTGGATGACGGTCTC 916

RESULT 13
AF125374/c
LOCUS
DEFINITION Ovis aries pyridoxal kinase mRNA, complete cds.
ACCESSION AF125374


```

VERSION AF125374.1 GI:4959454
KEYWORDS
SOURCE Ovis aries (sheep)
ORGANISM Ovis aries
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
          Bovidae; Caprinae; Ovis.
REFERENCE 1 (bases 1 to 996)
AUTHORS Kwon,O.-S. and Lee,H.-S.
TITLE Direct Submission
JOURNAL Natural Sciences, Kyungpook National University, College of
Puk-gu, Taegu 702 - 701 Korea

FEATURES
source
1..996
   Location/Qualifiers
   /organism="Ovis aries"
   /mol_type="mRNA"
   /db_xref="taxon:9940"
   /tissue_type="liver"
10..903
   /function="required for synthesis of pyridoxal
5'-phosphate from vitamin B6"
   /codon_start=1
   /product="pyridoxal kinase"
   /protein_id="AAD34353.1"
   /db_xref="GI:4959455"
   /translation="MCGGTAESLCPGLMQRHGLTWSALPTTPPGYSHMKGVNLN
DDVQELDGLKLNHNQYDYLVTGYTRDKSFLANVYDVOELKQONPRLYVCDPVMG
DQRNGEAMVPDGLFVREKVPVADITFNQFAELLTKRKHISGEAELEVDML
HSMGPDVTWITSDLSPPQSNLYLVIGSGRRNPAGSVVMERIMDIRKVDVAVFGT
GDLFAAMLLAWTHKHPNNLKVACEKTVSTLHLVLTQITQCAKAQAGSGVPSMQLEL
RMVQSKKDIESPIVQATVL"

CDS
190..909
   /note="pyridoxine; vitamin B6"
   /codon_start=1
   /product="pyridoxal kinase"
   /protein_id="AAP73047.1"
   /db_xref="GI:32186842"
   /translation="MNKYDYVLVTGYTRDKSFLANVYDVOELKQONPRLYVCDPVLG
DKWDEGSMYVPEDLPVREKVPVADITFNQFAELLTKRKHISGEAEALRWMDML
HSMGPDVTWITSDLSPPQSNLYLVIGSGRRNPAGSVVMERIMDIRKVDVAVFGT
GDLFAAMLLAWTHKHPNNLKVACEKTVSTLHLVLTQITQCAKAQAGSGVPSMQLEL
RMVQSKKDIESPIVQATVL"

ORIGIN
Query Match 75.8%; Score 18.2; DB 9; Length 1069;
Best Local Similarity 87.0%; Pred. No. 7.6e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCGTGGCGCTGGATGACGGTCTC 24
|||||
DB 902 ACCGTGGCGCTGGATGACGGTCTC 880
|||||

RESULT 15
BC005825/c
LOCUS 1227 bp mRNA linear PRI 03-OCT-2003
DEFINITION Homo sapiens pyridoxal (pyridoxine, vitamin B6) kinase, mRNA (cdna
clone MGC:1687 IMAGE:2989843), complete cds.
ACCESSION BC005825
VERSION BC005825.2 GI:33873199
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
          Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1227)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
          Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
          Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
          Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
          Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
          Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
          Scheetz,T.E., Brownstein,M.J., Uudin,T.B., Toshitoki,S.,
          Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
          Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
          McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
          Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
          Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
          Fahey,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S.,
          Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
          Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
          Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
          Butlerfield,Y.S., Krzywinski,M.I., Skalska,U., Smallos,D.E.,
          Schnerch,A., Schein,J.E., Jones,S.J. and Marz,M.A.
          Generation and initial analysis of more than 15,000 full-length
          human and mouse cDNA sequences
          Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
          22388257
          12477932
          2 (bases 1 to 1227)
          Strausberg,R.
          Direct Submission
          Submitted (02-APR-2001) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
          NIH-MGC Project URL: http://mgc.nci.nih.gov
          On Aug 19, 2003 this sequence version replaced gi:13543316.
          Contact: MGC help desk
          Email: gcaps-re@mail.nih.gov
          Tissue Procurement: DCTD/DTIP
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
          DNA Sequencing by: Institute for Systems Biology

JOURNAL TITLE
JOURNAL MEDLINE
PUBMED 12477932
REFERENCE 2 (bases 1 to 1227)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (02-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT On Aug 19, 2003 this sequence version replaced gi:13543316.
          Contact: MGC help desk
          Email: gcaps-re@mail.nih.gov
          Tissue Procurement: DCTD/DTIP
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
          DNA Sequencing by: Institute for Systems Biology

```


http://www.systemsbio.org
Contact: amadan@systemsbio.org
Anup Madan, Jessica Rahey, Erin Helton, Mark Kettman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 3 Row: i Column: 20
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18644884.

FEATURES
source
1. .1227
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:1687 IMAGE:2989843"
/tissue_type="Ovary, adenocarcinoma"
/clone_lib="NIH_MGC_9"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1. .1227
/gene="PDXK"
/note="synonyms: PKH, PNK"
/db_xref="LocusID:8566"
/db_xref="MIM:179020"
199. .1053
/codon_start=1
/product="PDXK protein"
/protein_id="AAH05825.1"
/db_xref="GI:1354317"
/db_xref="LocusID:8566"
/translation="MEEEKRVLSIQSHVIRGVGNRAATFPQLVGLFEIDAVNSVQFS
NHTGYAHWKGQVLNDELQELYEGLRNNNNKYDYLTVCDPVLGDKWDGEGSMYVE
DLFPVYKEKVPVLAIIITNPFAELLSGRKHSBEALRVMDMLHSGPDVTWITSS
DLSPQGSNYLIVLGSQRRNPAGSVMERIRMDIRKVDVFGTGDLFAMLLAWTH
KHPNLLKVAACEKTVSTLHHVLQRTIQAKAQAGEGVPSFMQLELRNVQSKRDIEDPE
IVVQATVL"
214. .1050
misc_feature
/note="PdxK; Region: Pyridoxal/pyridoxine/pyridoxamine
kinase [Coenzyme metabolism]"
/db_xref="CDD:COG2240"

ORIGIN
Query Match 75.8%; Score 18.2; DB 9; Length 1227;
Best Local Similarity 87.0%; Pred.No. 7.2e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 2 ACCGTGGGCTGGATGACGCTC 24
||||| ||||| ||||| |||||
Db 1046 ACCGTGCCTGGACGACGATC 1024

Search completed: April 29, 2004, 06:01:27
Job time : 367.465 secs

This Page Blank (uspto)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:01:30 ; Search time 1892.42 Seconds
(without alignments)
378.717 Million cell updates/sec

Title: US-10-624-714-2

Perfect score: 24

Sequence: 1 gaccctgggctggatgacgtctc 24

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_estc:*

9: gb_esti:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_man:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gssI:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	19.8	82.5	965	29	CNS01SW9
2	19.2	80.0	148	9	AA062466
3	19.2	80.0	203	29	AG024372
c 4	19.2	80.0	337	9	AU178867

RESULT 1

CNS01SW9/c

LOCUS

DEFINITION

Accession

Version

Keywords

Source

Organism

Reference

Authors

Title

Journal

Medline

PubMed

Reference

Authors

5	19.2	80.0	809	29	CG128545
6	19.2	80.0	886	29	CG128544
c 7	19.2	80.0	940	10	BF258586
8	18.8	78.3	625	29	CD573428
c 9	18.4	76.7	1175	14	CD503465
10	18.2	75.8	115	10	AW243324
c 11	18.2	75.8	235	12	BM466440
12	18.2	75.8	241	9	AA642237
13	18.2	75.8	248	13	EX344668
c 14	18.2	75.8	250	13	BU532437
c 15	18.2	75.8	254	13	BU590227
c 16	18.2	75.8	264	14	CB050012
c 17	18.2	75.8	264	14	CB050013
18	18.2	75.8	269	10	BE938100
19	18.2	75.8	271	14	CA488935
20	18.2	75.8	272	9	AI302947
21	18.2	75.8	284	9	AI273138
c 22	18.2	75.8	285	12	BG251797
c 23	18.2	75.8	286	13	BU567412
c 24	18.2	75.8	288	13	BU589006
25	18.2	75.8	300	9	AA158035
26	18.2	75.8	303	9	AI370358
c 27	18.2	75.8	304	13	BU860955
28	18.2	75.8	307	9	AA931828
29	18.2	75.8	309	9	AI025078
30	18.2	75.8	318	9	AI765949
31	18.2	75.8	319	9	AI927640
32	18.2	75.8	326	10	BF594525
33	18.2	75.8	332	10	BE857501
34	18.2	75.8	343	9	AI198230
c 35	18.2	75.8	346	14	CD641155
36	18.2	75.8	347	9	AA430023
c 37	18.2	75.8	358	13	BQ221890
c 38	18.2	75.8	368	12	BG988557
39	18.2	75.8	376	9	AA159982
40	18.2	75.8	378	14	CD030047
41	18.2	75.8	391	10	BE939902
42	18.2	75.8	393	12	BG938813
c 43	18.2	75.8	396	9	AI563985
c 44	18.2	75.8	403	13	BU949492

ALIGNMENTS

CNS01SW9	965 bp	DNA	linear	GSS 01-SEP-2000
Tetraodon nigroviridis genome survey sequence PUC-Orl end of clone				
170N03 of library G from Tetraodon nigroviridis, genomic survey				
sequence.				
AL165762				
AL165762.1	GI:7803500			
GSS; genome survey sequence.				
Tetraodon nigroviridis				
Tetraodon nigroviridis				
Tetraodon nigroviridis				
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;				
Tetraodontidae; Tetraodontidae; Tetraodon.				
1				
Roest Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,				
Bernot, A., Fzames, C., Wincker, P., Brottier, P., Quetier, F.,				
Saurin, W. and Weissenbach, J.				
Estimate of human gene number provided by genome-wide analysis				
using Tetraodon nigroviridis DNA sequence				
Nat. Genet. 25 (2), 235-238 (2000)				
20296633				
10835645				
2				
Roest Crollius, H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,				
Fzames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,				

Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000).

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE

3 (bases 1 to 965)

Genoscope.
Direct Submission
AUTHORS

Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
source

1..965
Location/Qualifiers
/organism="Tetraodon nigroviridis"
/mol_type="Genomic DNA"
/db_xref="taxon:99883"
/clone_lib="170N03"
/clone_lib="G"
/note="Genoscope sequence ID : COAG170CG02SP1-end :
PUC-ORI"

ORIGIN

Query Match 82.5%; Score 19.8; DB 29; Length 965;
Best Local Similarity 91.3%; Pred. No. 2.3e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ACCGTGGCTGGATGACGGTCTC 24

Db 943 ACCGTGGCTGGATGACGGTCTC 921

RESULT 2

AA062466

LOCUS

AA062466 148 bp mRNA linear EST 03-FEB-1997
m168f07.x1 Stratagene mouse kidney (#937315) Mus musculus cDNA
Clone IMAGE:517189 5' similar to gb:U20225 Mus musculus
adenylosuccinate lyase (MOUSE); mRNA sequence.

ACCESSION
VERSION

AA062466

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 148)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

CONTACT: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LBNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:311037

Trace considered overall poor quality

Seq primer: -28ml3 rev1.Er from Amersham

High quality sequence stop: 1.

Location/Qualifiers

1..148

/organism="Mus musculus"

FEATURES

source

/mol_type="mRNA"
/db_xref="taxon:10090"
/clone_lib="Stratagene mouse kidney (#937315)"
/db_xref="IMAGE:517189"
/sex="females"
/tissue_type="kidney"
/dev_stage="4 weeks"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: kidney; Vector: pBluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGGCAGG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match 80.0%; Score 19.2; DB 9; Length 148;
Best Local Similarity 87.5%; Pred. No. 2.4e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GACCTGGCTGGATGACGGTCTC 24

Db 69 GAACCTGGCTGGATGACGGTCCC 92

RESULT 3

AG024372

LOCUS

DEFINITION

Oryza sativa (japonica cultivar-group) DNA, clone:ND4039 0.705 1A,
3' flanking sequence of Tos17 insertion in rice strain ND4039,
genomic survey sequence.

ACCESSION

AG024372

VERSION

AG024372.1

KEYWORDS

SOURCE

ORGANISM

Oryza sativa (japonica cultivar-group)

Oryza sativa (japonica cultivar-group)

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE

AUTHORS

Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K.,

Shinozuka, Y., Onosato, K. and Hirochika, H.

Target site specificity of the Tos17 retrotransposon shows a

preference for insertion within genes and against insertion in

retrotransposon-rich regions of the genome

Plant Cell 15 (8), 1771-1780 (2003)

22779046

PUBMED

REFERENCE

2 (bases 1 to 203)

AUTHORS

Miyao, A., Onosato, K. and Hirochika, H.

Direct Submission

Submitted (25-OCT-1999) Akio Miyao, National Institute of

Agrobiological Sciences, Molecular Genetics; 2-1-2, Kannondai,

tsukuba, Ibaraki 305-8602, Japan (E-mail:miyao@affrc.go.jp,

URL: http://tos.nias.affrc.go.jp/, Tel: 81-298-38-7020,

Fax: 81-298-38-7020)

Location/Qualifiers

1..203

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="genomic DNA"

/strain="ND4039"

/cultivar="Nipponbare"

/db_xref="taxon:39947"

/clone_lib="ND4039 0.705 1A"

/note="The 3' end of retrotransposon Tos17 was found

immediately upstream of this sequence."

/note="This insertion point is base 5396 in AF003893."

misc_feature

1

ORIGIN

Query Match 80.0%; Score 19.2; DB 29; Length 203;

```

Best Local Similarity 87.5%; Pred. No. 2.6e+03; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 3;

QY 1 GACCGTGGCTGGATGACGGTCTC 24
    |||||
Db 84 GACCGTGGCTGGATGACGGTCTC 107
    |||||

RESULT 4
AUI78867/c
LOCUS
DEFINITION
AUI78867 Medaka ovary cDNA library (Old) from HNI Oryzias latipes
cDNA clone OLD12.10c similar to pir|JC7079| homeobox protein ZHX1 -
human, mRNA sequence.
ACCESSION
AUI78867.1 GI:13427703
VERSION
AUI78867
KEYWORDS
SOURCE
ORGANISM
Oryzias latipes (Japanese medaka)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.
1 (bases 1 to 337)
Naruse, K., Mitani, H. and Tanaka, M.
Medaka EST Project in University of Tokyo (2001)
Unpublished (2001)
Contact: Kiyoshi Naruse
Department of Biological Sciences
Graduate School of Science, University of Tokyo
Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan
Tel: 81-3-5841-4443
Fax: 81-3-5841-4410
Email: naruse@biol.s.u-tokyo.ac.jp
This clone was isolated from Medaka ovary cDNA library (Old).
FEATURES
    source
    1..337
        /organism="Oryzias latipes"
        /mol_type="mRNA"
        /strain="HNI"
        /db_xref="taxon:9090"
        /clone="Old12.10c"
        /tissue_type="ovary"
        /dev_stage="adult"
        /clone_lib="Medaka ovary cDNA library (Old) from HNI"

ORIGIN
Query Match 80.0%; Score 19.2; DB 9; Length 337;
Best Local Similarity 87.5%; Pred. No. 2.9e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GACCGTGGCTGGATGACGGTCTC 24
    |||||
Db 111 GACCGTCTGCTGGAAGACGGTCTC 88
    |||||

RESULT 5
CG128545
LOCUS
DEFINITION
PUFZG80TD ZM 0.6_1.0 KB Zea mays genomic clone ZMMBta0767M15,
genomic survey sequence.
ACCESSION
CG128545
VERSION
CG128545.1 GI:34011982
KEYWORDS
GSS.
SOURCE
Zea mays
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 809)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
REFERENCE
    AUTHORS
    TITLE
    JOURNAL
    COMMENT

TITLE
JOURNAL
COMMENT

FEATURES
    source
    1..886
        /organism="Zea mays"
        /mol_type="genomic DNA"
        /strain="B73"
        /db_xref="taxon:4577"
        /clone="ZMMBta0767M15"
        /clone_lib="ZM 0.6_1.0 KB"
        /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
        Cot selected genomic DNA library"

ORIGIN
Query Match 80.0%; Score 19.2; DB 29; Length 886;

```

```

Maize Genomics Consortium
Unpublished (2003)
Other GSSs: PUFZG80TB
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
Location/Qualifiers
    1..809
        /organism="Zea mays"
        /mol_type="genomic DNA"
        /strain="B73"
        /db_xref="taxon:4577"
        /clone="ZMMBta0767M15"
        /clone_lib="ZM 0.6_1.0 KB"
        /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
        Cot selected genomic DNA library"

ORIGIN
Query Match 80.0%; Score 19.2; DB 29; Length 809;
Best Local Similarity 87.5%; Pred. No. 3.8e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GACCGTGGCTGGATGACGGTCTC 24
    |||||
Db 312 GACCGTGGCTGGATGACGGTCTC 335
    |||||

RESULT 6
CG128544/c
LOCUS
DEFINITION
PUFZG80TB ZM 0.6_1.0 KB Zea mays genomic clone ZMMBta0767M15,
genomic survey sequence.
ACCESSION
CG128544
VERSION
CG128544.1 GI:34011981
KEYWORDS
GSS.
SOURCE
Zea mays
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 886)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
REFERENCE
    AUTHORS
    TITLE
    JOURNAL
    COMMENT

TITLE
JOURNAL
COMMENT

FEATURES
    source
    1..886
        /organism="Zea mays"
        /mol_type="genomic DNA"
        /strain="B73"
        /db_xref="taxon:4577"
        /clone="ZMMBta0767M15"
        /clone_lib="ZM 0.6_1.0 KB"
        /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
        Cot selected genomic DNA library"

ORIGIN
Query Match 80.0%; Score 19.2; DB 29; Length 886;

```

Best Local Similarity 87.5%; Pred. No. 3.8e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GACCTGGCTGGATGACGGTCTC 24
|||||
Db 718 GCGCTGGCTGGATGACGGCATC 695

RESULT 7
LOCUS BF258586
DEFINITION HVSMF0016C06f Hordeum vulgare seedling root EST library HVcDNA0007 (Etolated and unstressed) Hordeum vulgare subsp. vulgare cDNA clone HVSMF0016C06f, mRNA sequence.

ACCESSION BF258586
VERSION BF258586.1 GI:11187699
SOURCE EST.
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 940)
AUTHORS Wing, R., Close, T.J., Kleinohs, A., Wise, R., Begum, D., Frisch, D., Yu, X., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton, R.D., Oates, R. and Main, D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex unstressed seedling root cDNA library; Unpublished (2001)
JOURNAL COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 177
Seq primer: AATTAACCCCTCACTAAGGG
High quality sequence start: 78
High quality sequence stop: 510.

FEATURES
source
1..940
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/submitter="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMF0016C06f"
/tissue_type="Seedling root"
/lab_hosts="TUC11"
/clone_lib="Hordeum vulgare seedling root EST library HVcDNA0007 (Etolated and unstressed)"
/notes="Vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI; Seeds were surface sterilized then germinated under axenic conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedling roots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 1 million pfu were in vivo excised to give paluescript SK(-) cDNA phagemids. These steps were performed in the TU Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also

see Close TJ, Wing R, Kleinohs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genomics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)

ORIGIN
Query Match 80.0%; Score 19.2; DB 10; Length 940;
Best Local Similarity 87.5%; Pred. No. 3.9e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GACCTGGCTGGATGACGGTCTC 24
|||||
Db 359 GACGGTGGCTGGATGACGGTCTC 382

RESULT 8
LOCUS CC573428
DEFINITION CH240_450G23.TARBAC13P2 CHORI-240 Bos taurus genomic clone CH240_450G23, genomic survey sequence.

ACCESSION CC573428
VERSION CC573428.1 GI:31913091
SOURCE GSS.
ORGANISM Bos taurus (cow)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE 1 (bases 1 to 625)
AUTHORS Holt, R., Stott, J., Yang, G., Barber, S., Smalls, D., Prabhu, A.-L., Tsai, M., Cloutier, A., Lee, D., Girn, N., Olson, T., Mayo, M., Butterfield, X., Kirkpatrick, R., Liu, J., Guin, R., Chan, A., Chiu, R., Mathewson, C., Wye, N., Masson, A., Brown-John, M., Jones, S., Schein, J., Marra, M., de Jong, P., Keale, J.W. and Kappes, S.M.
TITLE Bovine BAC End Sequences from Library CHORI-240, PLATES 399 to 478
JOURNAL COMMENT Unpublished (2003)
Other GSSs: CH240_450G23.T7
Contact: Rob Holt
Sequencing The British Columbia Cancer Agency Genome Science Centre
600 W. 10th Ave. Vancouver, British Columbia, Canada V5Z 4E6
Tel: 604-877-6085
Fax: 604-877-6276
Email: rholt@bccsc.ca
Clones are derived from the bovine BAC library CHORI-240 (<http://www.chori.org/bacpac/bovine240.htm>). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org). Clones may be purchased from BACPAC Resources (<http://www.chori.org/bacpac/ordering-information.htm>). This work was undertaken as part of the International Bovine BAC Mapping Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the British Columbia Genome Sciences Centre, Canada.
Plate: 450 row: G column: 23
Seq primer: SP6
Class: BAC ends

FEATURES
source
1..625
/organism="Bos taurus"
/mol_type="genomic DNA"
/strain="bred: Hereford"
/db_xref="taxon:9913"
/clone="CH240_450G23"
/sex="Male"
/cell_type="Blood"
/clone_lib="CHORI-240"
/note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI; Hereford bull Li Domino 99375; CHORI-240 Bovine BAC library (Male) produced by Pieter de Jong"

ORIGIN
Query Match 78.3%; Score 18.8; DB 29; Length 625;
Best Local Similarity 90.9%; Pred. No. 5e+03;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12143 row: i column: 05
High quality sequence stop: 234.
Location/Qualifiers

FEATURES

source
1. .235
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5503564"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 Kb. Library constructed by Life Technologies."

ORIGIN

Query Match 75.8%; Score 18.2; DB 12; Length 235;
Best Local Similarity 87.0%; Pred. No. 6.4e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy

2 ACCGTGGCGTGGATGACGGTCTC 24
||||| ||||| ||||| |||||

Db

131 ACCGTGGCGTGGACGACGATCTC 109

RESULT 12

AA642237 241 bp mRNA linear EST 27-OCT-1997
LOCUS n:81b02.61 NCI_CGAP_Pr24 Homo sapiens cDNA clone IMAGE:1174347 3',
DEFINITION mRNA sequence.
ACCESSION AA642237 GI:2567455
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 241)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
COMMENT Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Suzanne L. Topalian, M.D., Robert K. Bright,
Ph.D.
cDNA Library Preparation: Stratagene, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image.html
Seq primer: -40m13 fwd. Ef from Amersham
High quality sequence stop: 177.
Location/Qualifiers

FEATURES

source
1. .241
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1174347"
/tissue_type="invasive tumor (cell line)"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="NCI_CGAP_Pr24"

/note="Organ: prostate; Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Invasive prostate tumor cell line (HPV immortalized). 5' adaptor sequence: 5' GAATTCGCGACGAG 3', adaptor sequence: 5' CTCAGTTTTTTTTTTTTTTT 3', Average insert size: 1.0 kb."

ORIGIN

Query Match 75.8%; Score 18.2; DB 9; Length 241;
Best Local Similarity 87.0%; Pred. No. 6.5e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy

2 ACCGTGGCGTGGATGACGGTCTC 24
||||| ||||| ||||| |||||

Db

66 ACCGTGGCGTGGACGACGATCTC 88

RESULT 13

EX344668 248 bp mRNA linear EST 02-MAY-2003
LOCUS EX344668 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DL002YA22 3-PRIME, mRNA sequence.
ACCESSION EX344668
VERSION EX344668.1 GI:30347326
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 248)
Li.W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNAL
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 10357.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS1AL001ZC12NP1&cluster=10357.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS1AL001ZC12NP1.
Location/Qualifiers

FEATURES

source
1. .248
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL002YA22"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 75.8%; Score 18.2; DB 13; Length 248;
Best Local Similarity 87.0%; Pred. No. 6.5e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy

2 ACCGTGGCGTGGATGACGGTCTC 24
||||| ||||| ||||| |||||

Db

119 ACCGTGGCGTGGACGACGATCTC 141

RESULT 14

BUS32437/c 250 bp mRNA linear EST 13-SEP-2002
LOCUS BUS32437


```

DEFINITION AGENCOURT_10200696 NIH_MGC_126 Homo sapiens cDNA clone
ACCESSION IMAGE:6558050 5', mRNA sequence.
VERSION BU532437
KEYWORDS BU532437.1 GI:22842878
SOURCE EST.
ORGANISM Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 250)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
        Email: cgabbs@mail.nih.gov
        Tissue Procurement: NCI
        cDNA Library Preparation: Michael Brownstein Laboratory
        cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
        DNA Sequencing by: Agencourt Bioscience Corporation
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LLNL at:
        http://image.llnl.gov
        Plate: LLCM2727 row: b column: 02
        High quality sequence stop: 249.
FEATURES             source
     source
     1..250
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="IMAGE:6558050"
     /tissue_type="mixed (pool of 40 RNAs)"
     /lab_host="DH10B (Tl-phage-resistant)"
     /clone_lib="NIH_MGC_126"
     /note="Vector: pDNR-LIB; Site 1: SfiI (ggcattatggcc);
     Site 2: SfiI (ggccctctggcc); Double-stranded cDNA was
     prepared from a pool of 40 cell line polyA+ RNAs (bladder
     - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
     4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
     kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell -
     5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%,
     salivary gland - 1.3%, and skin - 2.3%). 5' and 3'
     adaptors were used in cloning as follows:
     5'-AAGCAGTGGTATCAACGAGCGGCGCATTCACGCCGGG-3' and
     5'-ATTCTAGAGCGGCGGCGGCATG-dt(30)NN-3'. Full-length
     enriched library was constructed using the Clontech
     Creator SMART kit and size-selected to contain the 0.5-1
     kb size fraction (other fractions present in NIH_MGC_127
     and NIH_MGC_128). Library created in the laboratory of T.
     Usdin, M.D., Ph.D. (NIMH, NIH). Note: this is a NIH_MGC
     Library."
ORIGIN
Query Match 75.8%; Score 18.2; DB 13; Length 250;
Best Local Similarity 87.0%; Pred. No. 6.5e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCGTGGCTGGATGACGGTCTC 24
    ||||| ||||| ||||| |||||
Db 97 ACCGTGGCTGGACGACGATCTC 75

Search completed: April 29, 2004, 11:36:46
Job time : 1896.42 secs

DEFINITION AGENCOURT_8902968 NIH_MGC_141 Homo sapiens cDNA clone IMAGE:6389333
ACCESSION BU590227
VERSION BU590227.1 GI:23241667
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 250)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
        Email: cgabbs@mail.nih.gov
        Tissue Procurement: NCI
        cDNA Library Preparation: Michael Brownstein Laboratory
        cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
        DNA Sequencing by: Agencourt Bioscience Corporation
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LLNL at:
        http://image.llnl.gov
        Plate: LLCM2727 row: b column: 02
        High quality sequence stop: 249.
FEATURES             source
     source
     1..250
     /organism="Homo sapiens"
     /mol_type="mRNA"
     /db_xref="taxon:9606"
     /clone="IMAGE:6558050"
     /tissue_type="mixed (pool of 40 RNAs)"
     /lab_host="DH10B (Tl-phage-resistant)"
     /clone_lib="NIH_MGC_126"
     /note="Vector: pDNR-LIB; Site 1: SfiI (ggcattatggcc);
     Site 2: SfiI (ggccctctggcc); Double-stranded cDNA was
     prepared from a pool of 40 cell line polyA+ RNAs (bladder
     - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
     4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
     kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell -
     5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%,
     salivary gland - 1.3%, and skin - 2.3%). 5' and 3'
     adaptors were used in cloning as follows:
     5'-AAGCAGTGGTATCAACGAGCGGCGCATTCACGCCGGG-3' and
     5'-ATTCTAGAGCGGCGGCGGCATG-dt(30)NN-3'. Full-length
     enriched library was constructed using the Clontech
     Creator SMART kit and size-selected to contain the 0.5-1
     kb size fraction (other fractions present in NIH_MGC_127
     and NIH_MGC_128). Library created in the laboratory of T.
     Usdin, M.D., Ph.D. (NIMH, NIH). Note: this is a NIH_MGC
     Library."
ORIGIN
Query Match 75.8%; Score 18.2; DB 13; Length 250;
Best Local Similarity 87.0%; Pred. No. 6.5e+03;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCGTGGCTGGATGACGGTCTC 24
    ||||| ||||| ||||| |||||
Db 97 ACCGTGGCTGGACGACGATCTC 75

Search completed: April 29, 2004, 11:36:46
Job time : 1896.42 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:13:15 ; Search time 42.3529 Seconds
(without alignments)
314.472 Million cell updates/sec

Title: US-10-624-714-2

Perfect score: 24

Sequence: 1 gaccgtggctggatgacggcttc 24

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq*
- 5: /cgn2_6/ptodata/2/ina/PTCTUS COMB.seq*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
C 1	24	100.0	4403765	3	US-09-103-840A-2
C 2	24	100.0	4411529	3	US-09-103-840A-1
C 3	17.8	74.2	333	4	US-08-651-155B-102
C 4	17.8	74.2	333	4	US-09-194-036B-102
C 5	17.6	73.3	1055	4	US-09-252-991A-13146
C 6	17.6	73.3	1149	4	US-09-252-991A-13308
C 7	17.6	73.3	2241	4	US-09-252-991A-12657
C 8	17.2	71.7	750	4	US-09-489-039A-5270
C 9	17.2	71.7	1212	4	US-09-489-039A-5241
C 10	17.2	71.7	2853	4	US-09-252-991A-9356
C 11	17.2	71.7	3417	4	US-09-252-991A-9381
C 12	16.8	70.0	888	3	US-09-188-930-13
C 13	16.8	70.0	888	4	US-09-312-283C-13
C 14	16.6	69.2	435	4	US-09-621-976-10741
C 15	16.6	69.2	1020	4	US-09-489-039A-2380
C 16	16.6	69.2	1234	4	US-09-023-655A-444
C 17	16.6	69.2	1542	3	US-09-385-028-16
C 18	16.6	69.2	1542	4	US-09-726-614-16
C 19	16.6	69.2	1542	4	US-09-385-040-16
C 20	16.6	69.2	1542	1	US-08-462-484-9
C 21	16.6	69.2	2925	1	US-08-441-147-9
C 22	16.6	69.2	2925	5	PTCTUS95-07536-9
C 23	16.6	69.2	6250	1	US-08-729-214-23
C 24	16.6	69.2	6250	3	US-09-028-934-23
C 25	16.6	69.2	8931	3	US-09-028-934-28
C 26	16.6	69.2	11604	3	US-09-385-028-13
C 27	16.6	69.2	11604	4	US-09-726-614-13

C 28	16.6	69.2	11604	4	US-09-385-040-13	Sequence 13, Appl
C 29	16.6	69.2	15079	3	US-09-385-028-1	Sequence 1, Appl
C 30	16.6	69.2	15079	4	US-09-726-614-1	Sequence 1, Appl
C 31	16.6	69.2	15120	4	US-09-385-040-1	Sequence 1, Appl
C 32	16.6	69.2	80161	3	US-09-036-987A-1	Sequence 1, Appl
C 33	16.6	69.2	80161	3	US-09-370-700-1	Sequence 1, Appl
C 34	16.6	69.2	80161	4	US-09-603-207-1	Sequence 1, Appl
C 35	16.2	67.5	1269	1	US-08-396-218-1	Sequence 1, Appl
C 36	16.2	67.5	1269	1	US-08-760-116-1	Sequence 1, Appl
C 37	16.2	67.5	2081	2	US-09-096-982-7	Sequence 7, Appl
C 38	16.2	67.5	2081	2	US-08-653-650A-7	Sequence 7, Appl
C 39	16.2	67.5	3013	2	US-09-096-982-6	Sequence 6, Appl
C 40	16.2	67.5	3013	2	US-08-653-650A-6	Sequence 6, Appl
C 41	16.2	67.5	3196	2	US-09-096-982-4	Sequence 4, Appl
C 42	16.2	67.5	3196	2	US-08-653-650A-4	Sequence 4, Appl
C 43	16.2	67.5	4810	4	US-09-596-824-5	Sequence 5, Appl
C 44	16.2	67.5	4810	4	US-09-885-329-5	Sequence 5, Appl
C 45	16.2	67.5	23673	4	US-09-773-816-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 100.0%; Score 24; DB 3; Length 4403765;

Best Local Similarity 100.0%; Pred. No. 0.14;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACCGTGGGCTGGATGACGGTCTC 24

DB 55534 GACCGTGGGCTGGATGACGGTCTC 55511

RESULT 2

US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1

Query Match      100.0%; Score 24; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACCGTGGGCTGGATGACGCTTC 24
   |||||
DB 55592 GACCGTGGGCTGGATGACGCTC 55569

RESULT 3
US-08-651-155B-102
; Sequence 102, Application US/08651155B
; Patent No. 6365401
; GENERAL INFORMATION:
; APPLICANT: Mahan Dr., Michael J.
; APPLICANT: Conner Mr., Christopher P.
; APPLICANT: Hiethoff Mr., Douglas M.
; TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION
; TITLE OF INVENTION: OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST
; TITLE OF INVENTION: INFECTION
; NUMBER OF SEQUENCES: 255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chrisman, Bynum & Johnson, P.C.
; STREET: 1900 Fifteenth Street
; CITY: Boulder
; STATE: CO
; COUNTRY: USA
; ZIP: 80302
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/651,155B
; FILING DATE: 17-May-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Petersen Mr., Steven C.
; REGISTRATION NUMBER: 36,238
; REFERENCE/DOCKET NUMBER: 17060.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/546-1300
; TELEFAX: 303/449-5426
; TELEX: ABA1475
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; US-08-651-155B-102

Query Match      74.2%; Score 17.8; DB 4; Length 333;
Best Local Similarity 90.5%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACCGTGGGCTGGATGACGCT 21
   |||||
DB 14 GACCGTTGGCTGGATGACGCT 34

US-09-194-036B-102
; Sequence 102, Application US/09194036B
; Patent No. 6548246
; GENERAL INFORMATION:
; APPLICANT: Mahan, Michael J.
; APPLICANT: Conner, Christopher P.
; APPLICANT: Hiethoff, Douglas M.
; TITLE OF INVENTION: METHOD AND PROBES FOR THE IDENTIFICATION
; TITLE OF INVENTION: OF MICROBIAL GENES SPECIFICALLY INDUCED DURING HOST
; TITLE OF INVENTION: INFECTION
; NUMBER OF SEQUENCES: 255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 755 Page Mill Road
; CITY: Mountain View
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/194,036B
; FILING DATE: 17-No. 6548246-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/08208
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: US 08/651,155
; FILING DATE: 1996-05-17
; ATTORNEY/AGENT INFORMATION:
; NAME: Shantanu Basu
; REGISTRATION NUMBER: 43,318
; REFERENCE/DOCKET NUMBER: 220002060601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 813-5995
; TELEFAX: (650) 494-0792
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 333 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; ORGANISM: DNA (other)
; SEQUENCE DESCRIPTION: SEQ ID NO: 102:
US-09-194-036B-102

Query Match      74.2%; Score 17.8; DB 4; Length 333;
Best Local Similarity 90.5%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACCGTGGGCTGGATGACGCT 21
   |||||
DB 14 GACCGTTGGCTGGATGACGCT 34

RESULT 5
US-09-252-991A-13146/c
; Sequence 13146, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
```

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9356
; LENGTH: 2853
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9356

Query Match 71.7%; Score 17.2; DB 4; Length 2853;
Best Local Similarity 86.4%; Pred. No. 90;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GACCGTGGCTGGATGACGGTC 22
Db 424 GACCGTGGCTGGTGTCCGGAC 403

RESULT 11
US-09-252-991A-9381/c
; Sequence 9381, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 9381
; LENGTH: 3417
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-9381

Query Match 71.7%; Score 17.2; DB 4; Length 3417;
Best Local Similarity 86.4%; Pred. No. 90;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GACCGTGGCTGGATGACGGTC 22
Db 1584 GACCGTGGCTGGTGTCCGGAC 1563

RESULT 12
US-09-188-930-13/c
; Sequence 13, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James Greg
; TITLE OF INVENTION: Compositions Isolated From Skin Cells
; TITLE OF INVENTION: and Methods For Their Use
; FILE REFERENCE: 11000.1011c1
; CURRENT APPLICATION NUMBER: US/09/188,930A
; CURRENT FILING DATE: 1998-11-09
; NUMBER OF SEQ ID NOS: 348
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 13
; LENGTH: 888
; TYPE: DNA
; ORGANISM: mouse
US-09-188-930-13

Query Match 70.0%; Score 16.8; DB 3; Length 888;
Best Local Similarity 90.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 CCGTGGCTGGATGACGGTC 22
Db 36 CCGTGGCTGGATGTCGGTC 17

RESULT 13
US-09-312-283C-13/c
; Sequence 13, Application US/09312283C
; Patent No. 6573095
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Strachan, Lorna
; APPLICANT: Sleeman, Matthew
; APPLICANT: Onrust, Rene
; APPLICANT: Murison, James G.
; APPLICANT: Kumble, Krishanand D. Isolated from Skin Cells
; TITLE OF INVENTION: Compositions Isolated from Skin Cells
; TITLE OF INVENTION: and Methods for their Use
; FILE REFERENCE: 11000.1011c2
; CURRENT APPLICATION NUMBER: US/09/312,283C
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 425
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 888
; TYPE: DNA
; ORGANISM: Mouse
US-09-312-283C-13

Query Match 70.0%; Score 16.8; DB 4; Length 888;
Best Local Similarity 90.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 CCGTGGCTGGATGACGGTC 22
Db 36 CCGTGGCTGGATGTCGGTC 17

RESULT 14
US-09-621-976-10741/c
; Sequence 10741, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 10741
; LENGTH: 415
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-10741

Query Match 69.2%; Score 16.6; DB 4; Length 415;
Best Local Similarity 82.6%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GACCGTGGCTGGATGACGGTCT 23
```

Db 182 GACCGGGAGCTGTAGGACGGTCT 160

RESULT 15
 US-09-489-039A-2380/c
 ; Sequence 2380, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT APPLICATION NUMBER: US/09/489,039A
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: US 60/117,747
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 2380
 ; LENGTH: 1020
 ; TYPE: DNA
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-2380

Query Match 69.2%; Score 16.6; DB 4; Length 1020;
 Best Local Similarity 82.6%; Pred. No. 1.6e+02;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GACCGTGGCTGTAGGACGGTCT 23
 Db 699 GACCGGGAGCTGTAGGACGGTCT 677

Search completed: April 29, 2004, 11:44:52
 Job time : 53.3529 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 06:08:26 ; Search time 448 Seconds
(without alignments)
242.328 Million cell updates/sec

Title: US-10-624-714-2

Perfect score: 24

Sequence: 1 gaccgtggctggaagcgtctc 24

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA:

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	24	100.0	2037	9	US-09-712-363-12
C 2	24	100.0	2037	13	US-10-282-122A-28150
C 3	24	100.0	2466	13	US-10-282-122A-26493
C 4	20.8	86.7	1839	13	US-10-282-122A-25571
C 5	18.4	76.7	2388	16	US-10-369-493-33447
C 6	18.2	75.8	669	16	US-10-369-493-31738
C 7	18.2	75.8	960	9	US-09-880-107-3451
C 8	18.2	75.8	960	13	US-10-342-887-942
C 9	18.2	75.8	960	13	US-10-173-118-942
C 10	18.2	75.8	2169	15	US-10-240-965-17
C 11	18	75.0	506	10	US-09-918-995-32885
C 12	17.8	74.2	569	15	US-10-029-386-9670
C 13	17.6	73.3	582	16	US-10-369-493-44424
C 14	17.6	73.3	692	13	US-10-425-114-30827

C 15	17.6	73.3	957	13	US-10-282-122A-31580	Sequence 31580, A
C 16	17.6	73.3	1620	15	US-10-156-761-4407	Sequence 4407, Ap
C 17	17.6	73.3	1908	13	US-10-282-122A-20321	Sequence 20321, A
C 18	17.6	73.3	2139	13	US-10-282-122A-13725	Sequence 13725, A
C 19	17.6	73.3	9025608	15	US-10-156-761-1	Sequence 1, Appl
C 20	17.4	72.5	537	13	US-10-027-632-280479	Sequence 280479, A
C 21	17.4	72.5	537	16	US-10-027-632-280479	Sequence 280479, A
C 22	17.2	71.7	624	9	US-09-974-300-7421	Sequence 7421, Ap
C 23	17.2	71.7	948	13	US-10-425-114-1784	Sequence 1784, Ap
C 24	17.2	71.7	1065	16	US-10-369-493-26381	Sequence 26381, A
C 25	17.2	71.7	1333	13	US-10-425-114-13479	Sequence 13479, A
C 26	17.2	71.7	1452	13	US-10-425-114-22181	Sequence 22181, A
C 27	17.2	71.7	1476	13	US-10-047-825-36	Sequence 36, Appl
C 28	17.2	71.7	1515	13	US-10-425-114-1871	Sequence 1871, Ap
C 29	17.2	71.7	1532	13	US-10-425-114-22753	Sequence 22753, A
C 30	17.2	71.7	1690	9	US-09-986-632-9	Sequence 9, Appl
C 31	17.2	71.7	1956	13	US-10-282-122A-37086	Sequence 37086, A
C 32	17.2	71.7	2699	16	US-10-133-937-30	Sequence 30, Appl
C 33	17.2	71.7	2699	16	US-10-159-563-30	Sequence 30, Appl
C 34	17.2	71.7	3741	13	US-10-282-122A-38700	Sequence 38700, A
C 35	17	70.8	446	13	US-10-027-632-281006	Sequence 281006, A
C 36	17	70.8	446	16	US-10-027-632-281006	Sequence 281006, A
C 37	16.8	70.0	193	15	US-10-029-386-15792	Sequence 15792, A
C 38	16.8	70.0	532	13	US-10-029-386-2092	Sequence 2092, Ap
C 39	16.8	70.0	612	16	US-10-027-632-196463	Sequence 196463, A
C 40	16.8	70.0	880	10	US-09-866-050A-421	Sequence 421, App
C 41	16.8	70.0	880	15	US-10-152-661-421	Sequence 421, App
C 42	16.8	70.0	880	10	US-09-866-050A-13	Sequence 13, Appl
C 43	16.8	70.0	888	15	US-10-152-661-13	Sequence 13, Appl
C 44	16.8	70.0	888	15	US-10-152-661-13	Sequence 865, App
C 45	16.8	70.0	1134	16	US-10-398-221-865	

ALIGNMENTS

RESULT 1
US-09-712-363-12/c
; Sequence 12, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712.363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 2037
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis

US-09-712-363-12

Query Match 100.0%; Score 24; DB 9; Length 2037;
Best Local Similarity 100.0%; Pred. No. 0.13; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0;

Qy 1 GACCGTGGGCTGGATGACGGTCTC 24
|||||
Db 1932 GACCGTGGGCTGGATGACGGTCTC 1909

RESULT 2

US-10-282-122A-28150/c
Sequence 28150, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 28150

LENGTH: 2037

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

US-10-282-122A-28150

Query Match 100.0%; Score 24; DB 13; Length 2037;
Best Local Similarity 100.0%; Pred. No. 0.13; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0;

Qy 1 GACCGTGGGCTGGATGACGGTCTC 24
|||||
Db 1932 GACCGTGGGCTGGATGACGGTCTC 1909

RESULT 3

US-10-282-122A-26493/c

Sequence 26493, Application US/10282122A

Publication No. US20040029129A1
GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 26493

LENGTH: 2456

TYPE: DNA

ORGANISM: Mycobacterium bovis

US-10-282-122A-26493

Query Match 100.0%; Score 24; DB 13; Length 2466;
Best Local Similarity 100.0%; Pred. No. 0.12; Indels 0; Gaps 0;
Matches 24; Conservative 0; Mismatches 0;

Qy 1 GACCGTGGGCTGGATGACGGTCTC 24
|||||
Db 2364 GACCGTGGGCTGGATGACGGTCTC 2341

RESULT 4

US-10-282-122A-25571/c

Sequence 25571, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.

```

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 25571
; TYPE: DNA
; ORGANISM: Mycobacterium avium
US-10-282-122A-25571

Query Match      86.7%; Score 20.8; DB 13; Length 1839;
Best Local Similarity 91.7%; Pred. No. 3.8;
Matches 22; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACCGTGGGCTGGATGACGGTCTC 24
    |||||
Db 1668 GATGCTGGGCTGGATGACGGTCTC 1645

RESULT 5
US-10-369-493-33447
; Sequence 33447, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 33447
; LENGTH: 2388
; TYPE: DNA
; ORGANISM: Desulfotobacterium hafnienae
US-10-369-493-33447

Query Match      76.7%; Score 18.4; DB 16; Length 2388;
Best Local Similarity 95.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACCGTGGGCTGGATGACGG 20
    |||||
Db 1014 GACCGTGGGCTGGATGACCG 1033
```

```

RESULT 6
US-10-369-493-31738/c
; Sequence 31738, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 31738
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Rhodobacter sphaeroides
US-10-369-493-31738

Query Match      75.8%; Score 18.2; DB 16; Length 669;
Best Local Similarity 87.0%; Pred. No. 65;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GACCGTGGGCTGGATGACGGTCT 23
    |||||
Db 243 GACCGAGGGCTGGCTGATGCTCT 221

RESULT 7
US-09-880-107-3451/c
; Sequence 3451, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 3451
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U89606
US-09-880-107-3451

Query Match      75.8%; Score 18.2; DB 9; Length 960;
Best Local Similarity 87.0%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCGTGGGCTGGATGACGGTCTC 24
    |||||
Db 938 ACCGTGGGCTGGACGACGATCTC 916

RESULT 8
US-10-342-887-942/c
```

```
; Sequence 942, Application US/10342897
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 942
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-942

Query Match 75.8%; Score 18.2; DB 13; Length 960;
Best Local Similarity 87.0%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCGTGGCTGGATGACGGTCTC 24
DB 938 ACCGTGGCTGGATGACGGTCTC 916

RESULT 9
US-10-172-118-942/c
; Sequence 942, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 942
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_003681
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-942

Query Match 75.8%; Score 18.2; DB 13; Length 960;
Best Local Similarity 87.0%; Pred. No. 63;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCGTGGCTGGATGACGGTCTC 24
DB 938 ACCGTGGCTGGATGACGGTCTC 916
```

```
RESULT 10
US-10-240-965-17/c
; Sequence 17, Application US/10240965
; Publication No. US20030165924A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: SHIFFMAN, Dov
; APPLICANT: SOMOGYI, Roland
; APPLICANT: LAMN, Richard M.
; APPLICANT: SEILHAMER, Jeffrey J.
; APPLICANT: PORTER, Gordon J.
; APPLICANT: MIKITA, Thomas
; APPLICANT: TAI, Julie
; TITLE OF INVENTION: GENES EXPRESSED IN FOAM CELL DIFFERENTIATION
; FILE REFERENCE: PA-0025 PCT
; CURRENT APPLICATION NUMBER: US/10/240,965
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: 60/195,106
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 276
; SOFTWARE: PERL Program
; SEQ ID NO 17
; LENGTH: 2169
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030165924A1 076460.2
US-10-240-965-17

Query Match 75.8%; Score 18.2; DB 15; Length 2169;
Best Local Similarity 87.0%; Pred. No. 59;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ACCGTGGCTGGATGACGGTCTC 24
DB 1092 ACCGTGGCTGGATGACGGTCTC 1070

RESULT 11
US-09-918-995-32885/c
; Sequence 32885, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: HYSEQ, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32885
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(506)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-32885

Query Match 75.0%; Score 18; DB 10; Length 506;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACCGTGGCTGGATGAC 18
DB 302 GACCGTGGCTGGATGAC 285
```

RESULT 12
US-10-029-386-9670/c
; Sequence 9670, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 9670
; LENGTH: 569
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR3.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3
; OTHER INFORMATION: NT HIT: G14733256, EVALUE 1.00e-129
; OTHER INFORMATION: EST HUMAN HIT: AL046768.1, EVALUE 1.00e-129
; OTHER INFORMATION: SWISSPROT HIT: Q04637, EVALUE 3.00e-08
US-10-029-386-9670

Query Match 74.2%; Score 17.8; DB 15; Length 569;
Best Local Similarity 90.3%; Pred. No. 1e-02;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GACCGTGGGCTGGATGACGGT 21
DB 50 GACCGTGGCGGGATGCCGT 30

RESULT 13
US-10-369-493-44424/c
; Sequence 44424, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 44424
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Rhodospseudomonas palustris
US-10-369-493-44424

Query Match 73.3%; Score 17.6; DB 16; Length 582;
Best Local Similarity 83.3%; Pred. No. 1.3e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GACCGTGGGCTGGATGACGGTCTC 24
DB 48 GACCGTGACCTCGATGTCGGTCTC 25

RESULT 14
US-10-425-114-30827/c
; Sequence 30827, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 30827
; LENGTH: 692
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73084H05_FLI
US-10-425-114-30827

Query Match 73.3%; Score 17.6; DB 13; Length 692;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GACCGTGGGCTGGATGACGGTCTC 24
DB 455 GACCGTGACTGGACGACGGTCTC 432

RESULT 15
US-10-282-122A-31580/c
; Sequence 31580, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09

```

; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31580
; LENGTH: 957
; TYPE: DNA
; ORGANISM: Pseudomonas putida
US-10-282-122A-31580

```

```

Query Match      73.3%; Score 17.6; DB 13; Length 957;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy      1 GACCGTGGGCTGGATGACGGTCTC 24
Db      936 GACCGGGTCTGGATGACGTGTGC 913

```

```

Search completed: April 29, 2004, 20:43:48
Job time : 453 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:23:09 ; Search time 214.627 Seconds
(without alignments)
514.627 Million cell updates/sec

Title: US-10-624-714-3

Perfect score: 26
Sequence: 1 gatgacggatctcgggcggaac 26

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 674726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	26	100.0	959	2 AAX34004	Aax34004 Mycobacte
2	26	100.0	960	2 AAX34003	Aax34003 Mycobacte
3	26	100.0	961	2 AAX34002	Aax34002 Mycobacte
4	26	100.0	1063	2 AAX34007	Aax34007 Mycobacte
5	26	100.0	1243	2 AAX34001	Aax34001 Mycobacte
6	26	100.0	110000	4 AAI99682_02	Continuation (3 of
7	26	100.0	110000	4 AAI99683_02	Continuation (3 of
8	24	92.3	498	2 AAX34006	Aax34006 Mycobacte
9	19.6	75.4	2476	6 ABQ44291	Abq44291 Oligonuci
10	19.6	75.4	2476	6 ABQ44290	Abq44290 Oligonuci
11	19.6	75.4	110000	4 AAI99682_34	Continuation (35 o
12	19.6	75.4	110000	4 AAI99683_34	Continuation (35 o
13	19.2	73.8	722	7 ACA48882	ACA48882 Prokaryot
14	19.2	73.8	11768	4 ABL29425	ABl29425 Drosophil
15	19.2	73.8	15960	4 ABL29424	ABl29424 Drosophil
16	18.6	71.5	4817	6 ABA93122	ABa93122 Human KIM
17	18.6	71.5	8933	6 ABA93121	ABa93121 Human 5'
18	18.6	71.5	20387	2 AAV62159	AAv62159 HSV-2 str
19	18.6	71.5	26338	2 AAV62134	AAv62134 HSV-2 str
20	18.6	71.5	117213	2 AAV62176	AAv62176 HSV-2 str
21	18.6	71.5	154746	6 AAD25519	AAa25519 Human her
22	18	69.2	1393	5 AAH87709	AAh87709 Peppermin
23	18	69.2	1510	6 ABQ55048	ABq55048 Human ova

C	24	18	69.2	1515	4	AAF76608	Aaf76608 Peppermin
C	25	18	69.2	1665	2	AXX06563	Axx06563 (-)-limon
C	26	18	69.2	1665	4	AAF76607	Aaf76607 Peppermin
C	27	18	69.2	1665	4	AAF76665	Aaf76665 Computer-
C	28	18	69.2	1665	4	AAF76664	Aaf76664 Computer-
C	29	18	69.2	1701	3	AAC44498	Aac44498 Zea mays
C	30	18	69.2	41907	8	ADA37416	Ada37416 Human rDN
C	31	18	69.2	42998	7	ADA14747	Ada14747 Human rib
C	32	18	69.2	42999	6	ABS65032	Abs65032 Invertebr
C	33	18	69.2	42999	7	ABX11086	Abx11086 Human rib
C	34	18	69.2	42999	9	AAD61411	Aad61411 Human rib
C	35	17.6	67.7	386	4	ABA08405	AbA08405 Human mem
C	36	17.6	67.7	574	7	ACA36173	Aca36173 Prokaryot
C	37	17.6	67.7	1007	7	ACA31350	Aca31350 Prokaryot
C	38	17.6	67.7	1014	4	AAS56045	Aas56045 Salmonell
C	39	17.6	67.7	1014	4	AAS52570	Aas52570 E. coli D
C	40	17.6	67.7	1014	7	ACA51431	Aca51431 Prokaryot
C	41	17.6	67.7	1014	7	ACA32641	Aca32641 Prokaryot
C	42	17.6	67.7	2392	4	AAS15732	Aas15732 DNA encod
C	43	17.6	67.7	3441	6	AAD23842	Aad23842 Human pro
C	44	17.6	67.7	3772	3	AAZ64961	Aaz64961 Membrane-
C	45	17.6	67.7	3772	4	AAS21439	Aas21439 Human CDN

ALIGNMENTS

RESULT 1
AAX34004
ID AAX34004 standard; DNA; 959 BP.
XX
AC AAX34004;
XX
DT 06-JUL-1999 (first entry)
XX
DE Mycobacterium species nucleic acid sequence 1C'.
XX
KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
XX hybridisation; detection; vaccine; immunisation; infection; ss.
XX
OS Mycobacterium sp.
XX
PN WO9909186-A2.
XX
PD 25-FEB-1999.
XX
PF 14-AUG-1998; 98WO-FR001813.
XX
PR 14-AUG-1997; 97FR-00010404.
XX
PR 11-SEP-1997; 97FR-00011325.
XX
PA (INSP) INST PASTEUR.
XX
PI Gicquel B, Portnoie D, Lim E, Pellicic V, Guigueno A;
XX Goguet De La Salmoniere Y;
XX
DR WPI; 1999-181045/15.
XX P-PSDB; AAY04749.
XX
PT Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in infection-associated
PT protein expression.
XX
PS Claim 22; Fig 1C'; 309pp; French.
XX
CC Sequences AAX34001-X34252 represent nucleic acids encoding secreted
CC proteins from various Mycobacterium species microorganisms. The
CC nucleotide sequences can be used as primers and probes for methods for
CC detecting and identifying mycobacteria, especially belonging to the M.
CC tuberculosis complex. The encoded proteins can be used in vaccines for
CC immunisation against a bacterial or viral infection
XX
SQ Sequence 959 BP; 164 A; 313 C; 324 G; 158 T; 0 U; 0 Other;

Query Match 100.0%; Score 26; DB 2; Length 959;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGAAC 26
|||||
DB 570 GATGACGGATCGTCGGGGCGGGAAC 595

RESULT 2

AA34003
ID AAX34003 standard; DNA; 960 BP.

AC AAX34003;

DT 06-JUL-1999 (first entry)

XX Mycobacterium species nucleic acid sequence 1B'.

XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection; ss.
XX

OS Mycobacterium sp.

XX WO9909186-A2.

XX 25-FEB-1999.

XX 14-AUG-1998; 98WO-FR001813.

XX 14-AUG-1997; 97FR-00010404.

XX 11-SEP-1997; 97FR-00011325.

XX (INSP) INST PASTEUR.

XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;

PI Goguet De La Salmoniere Y;

XX WPI: 1999-181045/15.

DR P-PSDB; AAY04748.

XX Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in infection-associated
protein expression.

PS Claim 22; Fig 1B'; 309pp; French.

XX Sequences AAX34001-X34252 represent nucleic acids encoding secreted
CC proteins from various Mycobacterium species microorganisms. The
CC nucleotide sequences can be used as primers and probes for methods for
CC detecting and identifying mycobacteria, especially belonging to the M.
CC tuberculosis complex. The encoded proteins can be used in vaccines for
CC immunisation against a bacterial or viral infection

XX Sequence 960 BP; 165 A; 313 C; 324 G; 158 T; 0 U; 0 Other;

QY Query Match 100.0%; Score 26; DB 2; Length 960;

Best Local Similarity 100.0%; Pred. No. 0.12;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGAAC 26

|||||
DB 571 GATGACGGATCGTCGGGGCGGGAAC 595

RESULT 3

AA34002

ID AAX34002 standard; DNA; 961 BP.

XX AAX34002;

DT 06-JUL-1999 (first entry)

XX Mycobacterium species nucleic acid sequence 1A'.

XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection; ss.
XX

OS Mycobacterium sp.

XX WO9909186-A2.

XX 25-FEB-1999.

XX 14-AUG-1998; 98WO-FR001813.

XX 14-AUG-1997; 97FR-00010404.

XX 11-SEP-1997; 97FR-00011325.

XX (INSP) INST PASTEUR.

XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;

PI Goguet De La Salmoniere Y;

XX WPI: 1999-181045/15.

DR P-PSDB; AAY04747.

XX Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in infection-associated
protein expression.

PS Claim 22; Fig 1A'; 309pp; French.

XX Sequences AAX34001-X34252 represent nucleic acids encoding secreted
CC proteins from various Mycobacterium species microorganisms. The
CC nucleotide sequences can be used as primers and probes for methods for
CC detecting and identifying mycobacteria, especially belonging to the M.
CC tuberculosis complex. The encoded proteins can be used in vaccines for
CC immunisation against a bacterial or viral infection

XX Sequence 961 BP; 165 A; 313 C; 325 G; 158 T; 0 U; 0 Other;

QY Query Match 100.0%; Score 26; DB 2; Length 961;

Best Local Similarity 100.0%; Pred. No. 0.12;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGAAC 26

|||||
DB 572 GATGACGGATCGTCGGGGCGGGAAC 597

RESULT 4

AA34007

ID AAX34007 standard; DNA; 1063 BP.

XX AAX34007;

DT 06-JUL-1999 (first entry)

XX Mycobacterium species nucleic acid sequence 2.

XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection; ss.
XX

OS Mycobacterium sp.

XX WO9909186-A2.

XX 25-FEB-1999.

XX 14-AUG-1998; 98WO-FR001813.

XX 14-AUG-1997; 97FR-00010404.

XX 11-SEP-1997; 97FR-00011325.

PA (INSP) INST PASTEUR.
 XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
 PI Goguet De La Salmoniere Y;
 XX WPI; 1999-181045/15.
 DR P-PSDB; AAY04752.
 XX
 DR Mycobacterial DNA vectors containing reporter constructs - for
 XX identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 PT
 XX Claim 22; Fig 2; 309pp; French.
 PS
 XX Sequences AAX34001-X34252 represent nucleic acids encoding secreted
 CC proteins from various Mycobacterium species microorganisms. The
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection
 XX
 SQ Sequence 1063 BP; 173 A; 343 C; 374 G; 173 T; 0 U; 0 Other;
 Query Match 100.0%; Score 26; DB 2; Length 1063;
 Best Local Similarity 100.0%; Pred. No. 0.12;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GATGACGGATCGTCGGGGCGGGAAC 26
 DB 362 GATGACGGATCGTCGGGGCGGGAAC 387
 RESULT 5
 ID AAX34001
 AC AAX34001 standard; DNA; 1243 BP.
 XX
 AC AAX34001;
 XX
 DT 06-JUL-1999 (first entry)
 XX
 DE Mycobacterium species nucleic acid sequence 1.
 XX
 DE Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection; ss.
 XX
 OS Mycobacterium sp.
 XX
 PN WO9909186-A2.
 XX
 PD 25-FEB-1999.
 XX
 PF 14-AUG-1998; 98WO-FR001813.
 XX
 PR 14-AUG-1997; 97FR-00010404.
 PR 11-SEP-1997; 97FR-00011325.
 XX
 XX (INSP) INST PASTEUR.
 XX
 XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
 PI Goguet De La Salmoniere Y;
 XX WPI; 1999-181045/15.
 DR P-PSDB; AAY04752.
 XX
 DR Mycobacterial DNA vectors containing reporter constructs - for
 XX identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 PT
 XX Claim 22; Fig 1; 309pp; French.
 PS
 XX Sequences AAX34001-X34252 represent nucleic acids encoding secreted
 CC proteins from various Mycobacterium species microorganisms. The
 CC nucleotide sequences can be used as primers and probes for methods for

CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection
 XX
 SQ Sequence 1243 BP; 218 A; 394 C; 422 G; 209 T; 0 U; 0 Other;
 Query Match 100.0%; Score 26; DB 2; Length 1243;
 Best Local Similarity 100.0%; Pred. No. 0.12;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GATGACGGATCGTCGGGGCGGGAAC 26
 DB 853 GATGACGGATCGTCGGGGCGGGAAC 878

RESULT 6

AAX199682_02

Continuation (3 of 45) of AAI199682 from base 200001 (Mycobacterium tuberculosis strain H:
 WP Sequence split into 45 fragments LOCUS AAI199682 Accession AAI199682

WP	Fragment Name	Begin	End
WP	AAI199682_00	1	110000
WP	AAI199682_01	100001	210000
WP	AAI199682_02	200001	310000
WP	AAI199682_03	300001	410000
WP	AAI199682_04	400001	510000
WP	AAI199682_05	500001	610000
WP	AAI199682_06	600001	710000
WP	AAI199682_07	700001	810000
WP	AAI199682_08	800001	910000
WP	AAI199682_09	900001	1010000
WP	AAI199682_10	1000001	1110000
WP	AAI199682_11	1100001	1210000
WP	AAI199682_12	1200001	1310000
WP	AAI199682_13	1300001	1410000
WP	AAI199682_14	1400001	1510000
WP	AAI199682_15	1500001	1610000
WP	AAI199682_16	1600001	1710000
WP	AAI199682_17	1700001	1810000
WP	AAI199682_18	1800001	1910000
WP	AAI199682_19	1900001	2010000
WP	AAI199682_20	2000001	2110000
WP	AAI199682_21	2100001	2210000
WP	AAI199682_22	2200001	2310000
WP	AAI199682_23	2300001	2410000
WP	AAI199682_24	2400001	2510000
WP	AAI199682_25	2500001	2610000
WP	AAI199682_26	2600001	2710000
WP	AAI199682_27	2700001	2810000
WP	AAI199682_28	2800001	2910000
WP	AAI199682_29	2900001	3010000
WP	AAI199682_30	3000001	3110000
WP	AAI199682_31	3100001	3210000
WP	AAI199682_32	3200001	3310000
WP	AAI199682_33	3300001	3410000
WP	AAI199682_34	3400001	3510000
WP	AAI199682_35	3500001	3610000
WP	AAI199682_36	3600001	3710000
WP	AAI199682_37	3700001	3810000
WP	AAI199682_38	3800001	3910000
WP	AAI199682_39	3900001	4010000
WP	AAI199682_40	4000001	4110000
WP	AAI199682_41	4100001	4210000
WP	AAI199682_42	4200001	4310000
WP	AAI199682_43	4300001	4410000
WP	AAI199682_44	4400001	4411529

Query Match 100.0%; Score 26; DB 4; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 0.11;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGAAC 26
 DB 41423 GATGACGGATCGTCGGGGCGGGAAC 41448

RESULT 7

AAI99683_02
Continuation (3 of 44) of AAI99683 from base 200001 (Mycobacterium tuberculosis strain H
WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683
WP Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
WP AAI99683_19 1900001 2010000
WP AAI99683_20 2000001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000
WP AAI99683_28 2800001 2910000
WP AAI99683_29 2900001 3010000
WP AAI99683_30 3000001 3110000
WP AAI99683_31 3100001 3210000
WP AAI99683_32 3200001 3310000
WP AAI99683_33 3300001 3410000
WP AAI99683_34 3400001 3510000
WP AAI99683_35 3500001 3610000
WP AAI99683_36 3600001 3710000
WP AAI99683_37 3700001 3810000
WP AAI99683_38 3800001 3910000
WP AAI99683_39 3900001 4010000
WP AAI99683_40 4000001 4110000
WP AAI99683_41 4100001 4210000
WP AAI99683_42 4200001 4310000
WP AAI99683_43 4300001 4403765

Query Match 100.0%; Score 26; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 GATGCGATCGTCGGGGCGGGAAC 26
|||||
Db 41534 GATGCGATCGTCGGGGCGGGAAC 41559

RESULT 8

AAI99683_02
Continuation (3 of 44) of AAI99683 from base 200001 (Mycobacterium tuberculosis strain H
WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683
WP Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
WP AAI99683_19 1900001 2010000
WP AAI99683_20 2000001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000
WP AAI99683_28 2800001 2910000
WP AAI99683_29 2900001 3010000
WP AAI99683_30 3000001 3110000
WP AAI99683_31 3100001 3210000
WP AAI99683_32 3200001 3310000
WP AAI99683_33 3300001 3410000
WP AAI99683_34 3400001 3510000
WP AAI99683_35 3500001 3610000
WP AAI99683_36 3600001 3710000
WP AAI99683_37 3700001 3810000
WP AAI99683_38 3800001 3910000
WP AAI99683_39 3900001 4010000
WP AAI99683_40 4000001 4110000
WP AAI99683_41 4100001 4210000
WP AAI99683_42 4200001 4310000
WP AAI99683_43 4300001 4403765

Query Match 100.0%; Score 26; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 1 GATGCGATCGTCGGGGCGGGAAC 26
|||||
Db 41534 GATGCGATCGTCGGGGCGGGAAC 41559

DE Mycobacterium species nucleic acid sequence 1F.

KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection; ss.

XX Mycobacterium sp.
OS WO9909186-A2.
PN 25-FEB-1999.
PD 14-AUG-1998; 98WO-FR001813.
PF 14-AUG-1997; 97FR-00010404.
PR 11-SEP-1997; 97FR-00011325.
XX (INSP) INST PASTEUR.
XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A,
PI Goguet De La Salmoniere Y,
XX WPI; 1999-181045/15.
DR P-PSDB; AAY04751.
XX Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in infection-associated
PT protein expression.
XX Claim 22; Fig 1F; 309pp; French.
XX Sequences AAX34001-X34252 represent nucleic acids encoding secreted
CC proteins from various Mycobacterium species microorganisms. The
CC nucleotide sequences can be used as primers and probes for methods for
CC detecting and identifying mycobacteria, especially belonging to the M.
CC tuberculosis complex. The encoded proteins can be used in vaccines for
CC immunisation against a bacterial or viral infection
XX SQ Sequence 498 BP; 87 A; 161 C; 173 G; 77 T; 0 U; 0 Other;
Query Match 92.3%; Score 24; DB 2; Length 498;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 3 TGACGGATCGTCGGGGCGGGAAC 26
|||||
Db 1 TGACGGATCGTCGGGGCGGGAAC 24
RESULT 9
ABQ44291/c
ID ABQ44291 standard; DNA; 2476 BP.
XX AC ABQ44291;
XX DT 12-JUL-2002 (first entry)
XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 30892.
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX OS Homo sapiens.
XX WO200218632-A2.
XX PD 07-MAR-2002.
XX PF 01-SEP-2001; 2001WO-EP010074.
XX PR 01-SEP-2000; 2000DE-01043826.
XX PR 05-SEP-2000; 2000DE-01044543.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful for

XX diagnosis and prognosis, comprises selective hybridization of amplicons

XX from chemically treated DNA.

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a

XX genomic sample of DNA. The sample is treated chemically to convert

XX cytosine (C) but not methylated C, to uracil, then part of the genomic

XX DNA that contains the target C is amplified to form a labeled amplicon.

XX The amplicon is hybridised to two classes, each with at least one member,

XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the

XX degree of hybridisation to both classes is determined from the label on

XX the amplicon. From the ratio of labels hybridised to the two classes of

XX oligomers, the degree of methylation is calculated. The method is used;

XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs

XX and of a wide range of diseases, e.g. cancer, disorders of the central

XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,

XX particularly by detecting mutations or single nucleotide polymorphisms

XX (SNP's); and (ii) for differentiation of cell or tissue types and for

XX investigating cell differentiation. The method allows the methylation

XX status of many C residues to be determined simultaneously. ABQ13410-

XX ABQ54121 represent genomic DNA sequences used to illustrate the method

XX for determining the degree of cytosine methylation described in the

XX disclosure of the invention

XX Sequence 2476 BP; 817 A; 881 C; 341 G; 436 T; 0 U; 1 Other;

XX Query Match 75.4%; Score 19.6; DB 6; Length 2476;

XX Best Local Similarity 84.6%; Pred. No. 61;

XX Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATGACCGATCGTCGGGGCGGGGAAAC 26

DB 2076 GAGGAGGGGGCGTCGGGGCGGGGAAAC 2051

RESULT 10

ABQ44290

ID ABQ44290 standard; DNA; 2476 BP.

AC ABQ44290;

XX 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 30881.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

XX drug; side effect; cancer; central nervous system; cardiovascular;

XX gastrointestinal; respiratory system; single nucleotide polymorphism;

XX SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP010074.

XX 01-SEP-2000; 2000DE-01043826.

XX 05-SEP-2000; 2000DE-01044543.

XX (EPIC-) EPICENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful for

PT diagnosis and prognosis, comprises selective hybridization of amplicons

PT from chemically treated DNA.

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of

XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a

XX genomic sample of DNA. The sample is treated chemically to convert

XX cytosine (C) but not methylated C, to uracil, then part of the genomic

XX DNA that contains the target C is amplified to form a labeled amplicon.

XX The amplicon is hybridised to two classes, each with at least one member,

XX of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the

XX degree of hybridisation to both classes is determined from the label on

XX the amplicon. From the ratio of labels hybridised to the two classes of

XX oligomers, the degree of methylation is calculated. The method is used;

XX (i) for diagnosis and/or prognosis of side effects of therapeutic drugs

XX and of a wide range of diseases, e.g. cancer, disorders of the central

XX nervous, cardiovascular, gastrointestinal and respiratory systems etc.,

XX particularly by detecting mutations or single nucleotide polymorphisms

XX (SNP's); and (ii) for differentiation of cell or tissue types and for

XX investigating cell differentiation. The method allows the methylation

XX status of many C residues to be determined simultaneously. ABQ13410-

XX ABQ54121 represent genomic DNA sequences used to illustrate the method

XX for determining the degree of cytosine methylation described in the

XX disclosure of the invention

XX Sequence 2476 BP; 436 A; 341 C; 881 G; 817 T; 0 U; 1 Other;

XX Query Match 75.4%; Score 19.6; DB 6; Length 2476;

XX Best Local Similarity 84.6%; Pred. No. 61;

XX Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATGACCGATCGTCGGGGCGGGGAAAC 26

DB 401 GAGGAGGGGGCGTCGGGGCGGGGAAAC 426

RESULT 11

AAI99682_34/c

Continuation (35 of 45) of AAI99682 from base 3400001 (Mycobacterium tuberculosis strain

WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000

CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 722 BP; 144 A; 186 C; 235 G; 157 T; 0 U; 0 Other;

Query Match 73.8%; Score 19.2; DB 7; Length 722;
Best Local Similarity 87.5%; Pred. No. 91;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GATGACCGATCGTCGGGGCGGGA 24
DB 293 GATGACCGATCGTCGGGGCTGGA 316

RESULT 14
ABL29425/c
ID ABL29425 standard; DNA; 11768 BP.
XX AC ABL29425;
DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 39748.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Claim 1; SEQ ID NO 39748; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
XX CC ABBS72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 11768 BP; 3122 A; 2770 C; 3081 G; 2795 T; 0 U; 0 Other;

Query Match 73.8%; Score 19.2; DB 4; Length 11768;
Best Local Similarity 87.5%; Pred. No. 88;

Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 ATGACGGATCGTCGGGGCGGGA 25
DB 2882 ATGACGGATCGTCGGGGCGGGA 2859
RESULT 15
ABL29424/c
ID ABL29424 standard; DNA; 15960 BP.
XX AC ABL29424;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 39745.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX DR New isolated nucleic acid detection reagent for detecting 1000 or more
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX PT interactions.
XX PS Claim 1; SEQ ID NO 39745; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is
XX CC useful in developmental biology and in elucidating cell signalling and
XX CC cell-cell interactions in higher eukaryotes for the development of
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS57737-
XX CC ABBS72072). The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 15960 BP; 4507 A; 3458 C; 3783 G; 4212 T; 0 U; 0 Other;

Query Match 73.8%; Score 19.2; DB 4; Length 15960;
Best Local Similarity 87.5%; Pred. No. 87;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 2 ATGACGGATCGTCGGGGCGGGA 25
DB 5206 ATGACGGATCGTCGGGGCGGGA 5183
Search completed: April 29, 2004, 04:56:54
Job time : 216.727 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:13:15 ; Search time 45.8824 Seconds
(without alignments)
314.472 Million cell updates/sec

Title: US-10-624-714-3

Perfect score: 26
Sequence: 1 gatgacggatctcggggcggaac 26

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/8A_COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/8B_COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	26	100.0	4403765	3 US-09-103-840A-2	Sequence 2, Appli
2	26	100.0	4411529	3 US-09-103-840A-1	Sequence 1, Appli
C 3	19.6	75.4	4403765	3 US-09-103-840A-2	Sequence 2, Appli
C 4	19.6	75.4	4411529	3 US-09-103-840A-1	Sequence 1, Appli
C 5	18.6	71.5	129	3 US-08-952-973-16	Sequence 16, Appli
C 6	18	69.2	1515	3 US-09-292-768-5	Sequence 5, Appli
C 7	18	69.2	1655	3 US-08-881-784-8	Sequence 8, Appli
C 8	18	69.2	1655	3 US-09-292-768-3	Sequence 3, Appli
C 9	18	69.2	1665	3 US-09-292-768-67	Sequence 67, Appli
C 10	18	69.2	1665	3 US-09-292-768-69	Sequence 69, Appli
C 11	17.6	67.7	1032	4 US-09-489-039A-2050	Sequence 2050, Ap
C 12	17.6	67.7	3470	4 US-09-976-594-345	Sequence 345, App
C 13	17.6	67.7	8378	5 PCT-US91-09055-1	Sequence 1, Appli
C 14	17.4	66.9	792	4 US-09-489-039A-1358	Sequence 1358, Ap
C 15	17.4	66.9	112	4 US-09-489-039A-4710	Sequence 4710, Ap
C 16	17.2	66.2	306	4 US-09-252-991A-6766	Sequence 6766, Ap
C 17	17.2	66.2	966	4 US-09-252-991A-6895	Sequence 6895, Ap
C 18	17	65.4	1434	4 US-09-489-039A-2415	Sequence 2415, Ap
C 19	17	65.4	11464	3 US-08-991-840A-2	Sequence 2, Appli
C 20	16.8	64.6	80161	3 US-09-036-987A-1	Sequence 1, Appli
C 21	16.8	64.6	80161	3 US-09-370-700-1	Sequence 1, Appli
C 22	16.8	64.6	80161	4 US-09-603-207-1	Sequence 1, Appli
C 23	16.6	63.8	279	4 US-09-313-294A-1175	Sequence 1175, Ap
C 24	16.6	63.8	481	4 US-09-116-492A-3	Sequence 3, Appli
C 25	16.6	63.8	524	4 US-09-116-492A-2	Sequence 2, Appli
C 26	16.6	63.8	855	4 US-09-116-492A-36	Sequence 36, Appli
C 27	16.6	63.8	1069	4 US-09-116-492A-37	Sequence 37, Appli

C	28	16.6	63.8	1083	4	US-09-252-991A-13913	Sequence 13913, A
	29	16.6	63.8	1104	4	US-09-894-844-3	Sequence 3, Appli
	30	16.6	63.8	1200	4	US-09-050-739-91	Sequence 91, Appli
	31	16.6	63.8	1263	4	US-09-252-931A-13538	Sequence 13538, A
	32	16.6	63.8	1277	4	US-09-116-492A-1	Sequence 1, Appli
	33	16.6	63.8	1282	4	US-09-116-492A-38	Sequence 38, Appli
	34	16.6	63.8	1386	4	US-09-489-039A-2452	Sequence 2452, Ap
	35	16.6	63.8	1616	3	US-08-818-112-112	Sequence 112, App
	36	16.6	63.8	1616	4	US-08-818-111-107	Sequence 107, App
	37	16.6	63.8	1616	4	US-09-056-556-112	Sequence 112, App
	38	16.6	63.8	1616	4	US-09-072-596-107	Sequence 107, App
	39	16.6	63.8	1616	4	US-09-072-567-112	Sequence 112, App
	40	16.6	63.8	1819	1	US-08-487-753-6	Sequence 6, Appli
	41	16.6	63.8	1819	2	US-08-480-065-6	Sequence 6, Appli
	42	16.6	63.8	1819	3	US-08-487-744-6	Sequence 6, Appli
	43	16.6	63.8	1819	5	PCT-US93-09167-6	Sequence 6, Appli
	44	16.6	63.8	2554	2	US-08-627-254C-28	Sequence 28, Appli
	45	16.6	63.8	16885	1	US-08-390-878-16	Sequence 16, Appli

ALIGNMENTS

RESULT 1

US-09-103-840A-2

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103.840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 100.0%; Score 26; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATGACGGATCTCGGGCGGGAAC 26

DB 241534 GATGACGGATCTCGGGCGGGAAC 241559

RESULT 2

US-09-103-840A-1

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103.840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      100.0%; Score 26; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGAAC 26
Db 241423 GATGACGGATCGTCGGGGCGGGAAC 241448

RESULT 3
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match      75.4%; Score 19.6; DB 3; Length 4403765;
Best Local Similarity 84.6%; Pred. No. 9.5;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGAAC 26
Db 3492898 GATGCGGCTCGTTGCGGGCGGGAAC 3492873

RESULT 4
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      75.4%; Score 19.6; DB 3; Length 4411529;
Best Local Similarity 84.6%; Pred. No. 9.5;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGAAC 26
Db 3497149 GATGCGGCTCGTTGCGGGCGGGAAC 3497124

RESULT 5
US-08-952-973-16/c
; Sequence 16, Application US/08952973
; Patent No. 6083717
; GENERAL INFORMATION:
; APPLICANT: Madzak, Catherine
; APPLICANT: Blanchin-Roland, Sylvie
; APPLICANT: Gaillardin, Claude
; TITLE OF INVENTION: UPSTREAM ACTIVATOR SEQUENCES AND
; TITLE OF INVENTION: RECOMBINANT PROMOTER SEQUENCES FUNCTIONAL IN YARROWIA AND
; TITLE OF INVENTION: VECTORS CONTAINING THEM
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/952,973
; FILING DATE: 06-JAN-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/IB96/00562
; FILING DATE: 06-JUN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95401322.3
; FILING DATE: 08-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Granados, Patricia D.
; REGISTRATION NUMBER: 33,683
; REFERENCE/DOCKET NUMBER: 065691/0129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 129 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-952-973-16

Query Match      71.5%; Score 18.6; DB 3; Length 129;
Best Local Similarity 84.0%; Pred. No. 36;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ATGACGGATCGTCGGGGCGGGAAC 26
Db 120 ATGCGGATCGAGTGGCGGGAAC 96

RESULT 6
US-09-292-768-5/c
```



```
; Sequence 5, Application US/09292768
; Patent No. 6194185
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lupien, Shari L
; APPLICANT: Karp, Frank
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF
; TITLE OF INVENTION: LIMONENE HYDROXYLASES
; FILE REFERENCE: wsur13463
; CURRENT APPLICATION NUMBER: US/09/292,768
; CURRENT FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: 08/881,784
; EARLIER FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1515
; TYPE: DNA
; ORGANISM: Mentha piperita
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1494)
US-09-292-768-5

Query Match      69.2%; Score 18; DB 3; Length 1515;
Best Local Similarity 80.8%; Pred. No. 59;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 GATGACGGATCGTCGGGGCGGGAAC 26
Db      1493 GAGGAGGATCGTAGGGTGTGGGAAC 1468

RESULT 7
US-08-881-784-8/c
; Sequence 8, Application US/0881874
; Patent No. 6083731
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B.
; APPLICANT: Lupien, Shari L.
; APPLICANT: Karp, Frank
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR
; TITLE OF INVENTION: THE PRODUCTION OF LIMONENE HYDROXYLASES
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; ADDRESSEE: PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/881,784
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; REFERENCE/DOCKET NUMBER: Wsur19777
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 224-0718
; TELEFAX: (206) 224-0779
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1665 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Mentha x piperita
; IMMEDIATE SOURCE:
; CLONE: pPM17
US-08-881-784-8

Query Match      69.2%; Score 18; DB 3; Length 1665;
Best Local Similarity 80.8%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 GATGACGGATCGTCGGGGCGGGAAC 26
Db      1517 GAGGAGGATCGTAGGGTGTGGGAAC 1492

RESULT 8
US-09-292-768-3/c
; Sequence 3, Application US/09292768
; Patent No. 6194185
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lupien, Shari L
; APPLICANT: Karp, Frank
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF
; TITLE OF INVENTION: LIMONENE HYDROXYLASES
; FILE REFERENCE: wsur13463
; CURRENT APPLICATION NUMBER: US/09/292,768
; CURRENT FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: 08/881,784
; EARLIER FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1665
; TYPE: DNA
; ORGANISM: Mentha piperita
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19)..(1518)
US-09-292-768-3

Query Match      69.2%; Score 18; DB 3; Length 1665;
Best Local Similarity 80.8%; Pred. No. 58;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY      1 GATGACGGATCGTCGGGGCGGGAAC 26
Db      1517 GAGGAGGATCGTAGGGTGTGGGAAC 1492

RESULT 9
US-09-292-768-67/c
; Sequence 67, Application US/09292768
; Patent No. 6194185
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lupien, Shari L
; APPLICANT: Karp, Frank
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF
; TITLE OF INVENTION: LIMONENE HYDROXYLASES
; FILE REFERENCE: wsur13463
; CURRENT APPLICATION NUMBER: US/09/292,768
; CURRENT FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: 08/881,784
; EARLIER FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 67
; LENGTH: 1665
; TYPE: DNA
; ORGANISM: Artificial Sequence
```

```

;
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: computer-generated nucleic acid sequence encoding
; OTHER INFORMATION: limonene-3-hydroxylase
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1665)
; OTHER INFORMATION: computer-generated nucleic acid sequence encoding
; OTHER INFORMATION: limonene-3-hydroxylase variant
;
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19)..(1518)
US-09-292-768-67

```

	Query Match	69.2%	Score 18	DB 3	Length 1665
	Best Local Similarity	80.8%	Prod. No. 58		
	Matches 21	Conservative 0	Mismatches 5	Indels 0	Gaps 0
QY	1	GATCACGGATCTCGGGGGCGGAAC	26		
Db	1517	GAGGAAGGATCGTAGGTTGTGGAAAC	1492		

```

RESULT 10
US-09-292-768-69/c
; Sequence 69, Application US/09292768
; Patent No. 6194185
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lupien, Shari L
; APPLICANT: Karp, Frank
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR THE PRODUCTION OF
; TITLE OF INVENTION: LIMONENE HYDROXYLASES
; FILE REFERENCE: wslr13463
; CURRENT APPLICATION NUMBER: US/09/292.768
; CURRENT FILING DATE: 1999-04-14
; EARLIER APPLICATION NUMBER: 08/881,784
; EARLIER FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 1665
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: computer-generated nucleic acid sequence encoding
; OTHER INFORMATION: limonene-3-hydroxylase
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1665)
; OTHER INFORMATION: computer-generated nucleic acid sequence encoding
; OTHER INFORMATION: limonene-3-hydroxylase protein variant
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (19)..(1518)
US-09-292-768-69

```

```

Query Match      69.2%  Score 18;  DB 3;  Length 1605;
Best Local Similarity 80.8%;  Pred.No. 58;
Matches 21;  Conservative 0;  Mismatches 5;  Indels 0;  Gaps 0;

QY      1  GATGACGGATCGTCGGGGGGGGGAAC 26
          |||||
Db      1517  GAGGAAGCATCGTAGGGTGTGGGAAC 1492
          |||||

```

RESULT 11
US-09-489-039A-2050
; Sequence 2050, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al

```

/ / TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
/ /
/ / TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
/ /
/ / FILE REFERENCE: 2709.2004001
/ /
/ / CURRENT APPLICATION NUMBER: US/09/489,039A
/ /
/ / CURRENT FILING DATE: 2000-01-27
/ /
/ / PRIOR APPLICATION NUMBER: US 60/117,747
/ /
/ / PRIOR FILING DATE: 1999-01-29
/ /
/ / NUMBER OF SEQ ID NOS: 14342
/ /
/ / SEQ ID NO 2050
/ /
/ / LENGTH: 1032
/ /
/ / TYPE: DNA
/ /
/ / ORGANISM: Klebsiella pneumoniae
/ /
/ / US-09-489-039A-2050

```

```

Query Match      67.7%; Score 17.6; DB 4; Length 1032;
Best Local Similarity 83.3%; Pred. No. 86;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1  GATGACGGATCTCGGGGGCGGA 24
          |||||
Db       618  GATGACCGACCGTCCGGGGCTGA 641

```

```

RESULT 12
US-09-976-594-345
; Sequence 345, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: FA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 345
; LENGTH: 3470
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6673549 1129157.1
US-09-976-594-345

```

Query Match	67.7%	Score 17.6	DB 4	Length 3470
Best Local Similarity	83.3%	Pred. NO. 83		
Matches	20	Conservative 0	Mismatches 4	Indels 0
Gaps	0			
QY	2	ATGACGGATCGTCGGGGCGGGAA	25	
Db	599	ATGACGGAGCGTGTGCGCGGGAA	622	

```

RESULT 13
PCT-US91-09055-1/c
; Sequence 1, Application PC/TUS9109055
; GENERAL INFORMATION:
; APPLICANT: Rohberg, Jonathan Marc and Artavanis-Tsakonas, Spyridon
; TITLE OF INVENTION: Purified SLIT protein and Sequence Elements Thereof
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yale University
; ADDRESSEE: Office of Cooperative Research
; STREET: 246 Church Street
; STREET: Suite 401
; CITY: New Haven
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06510
; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Diskette, 3.50 inch. 800 Kb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.5
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICANT: Klebsiella pneumoniae
PCT/US91/09055
FILING DATE: 19911127
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
PRIORITY APPLICATION NUMBER: US/07/624,135
FILING DATE: 7-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Barth, Richard J.
REGISTRATION NUMBER: 28,180
REFERENCE/DOCKET NUMBER: 900964/RSB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 972-1400
TELEFAX: (212) 370-1622
TELEX: 236268
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8378
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
HYPOTHETICAL: no
ANTI-SENSE: no
ORIGINAL SOURCE:
ORGANISM: Drosophila Melanogaster
DEVELOPMENTAL STAGE: embryonic and larval, germ-line
IMMEDIATE SOURCE:
LIBRARY: cDNA and Genomic
CLONE: be2.4, ka2.4, B52-2, B52-5, smart2-19
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 2R
MAP POSITION: 52D
UNITS: chromosome band
FEATURE:
NAME/KEY: 5' leader sequence
LOCATION: 1 to 314
IDENTIFICATION METHOD: experimental
NAME/KEY: Translated region
LOCATION: 315 to 4754
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: codes for slit protein
NAME/KEY: 3' untranslated region
LOCATION: 4755 to 8378
IDENTIFICATION METHOD: experimental
PCT-US91-09055-1

Query Match 67.7%; Score 17.6; DB 5; Length 8378;
Best Local Similarity 83.3%; Pred. No. 80;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGA 24
|||||
Db 7436 GATGACGGATCGTCGGGGCGGATGGA 7413

RESULT 14
US-09-489-039A-1358
; Sequence 1358, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1358
; LENGTH: 792
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1358

Query Match 66.9%; Score 17.4; DB 4; Length 792;
Best Local Similarity 94.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 GATCGTCGGGGCGGGAAC 26
|||||
Db 351 GATCGTCGGGGCGGGAAC 369

RESULT 15
US-09-489-039A-4710
; Sequence 4710, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 4710
; LENGTH: 1212
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4710

Query Match 66.9%; Score 17.4; DB 4; Length 1212;
Best Local Similarity 94.7%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 CGGATCGTCGGGGCGGGA 24
|||||
Db 43 CTGATCGTCGGGGCGGGA 61

Search completed: April 29, 2004, 11:45:14
Job time : 67.8824 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:25:04 ; Search time 392.345 Seconds
(without alignments)
2872.264 Million cell updates/sec

Title: US-10-624-714-3

Perfect score: 26
Sequence: 1 gatgacgatgctcggggcggaac 26

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl :

1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_hg_hum.*
31: em_hg_inv.*
32: em_hg_other.*
33: em_hg_mus.*
34: em_hg_pln.*
35: em_hg_rod.*
36: em_hg_nam.*
37: em_hg_vrt.*
38: em_sy.*
39: em_hgo_hum.*
40: em_hgo_mus.*
41: em_hgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	26	100.0	959	6	BD273780 Sequences
2	26	100.0	959	6	AX004906 Sequences
3	26	100.0	960	6	BD273779 Sequences
4	26	100.0	960	6	AX004895 Sequences
5	26	100.0	961	6	BD273778 Sequences
6	26	100.0	961	6	AX004889 Sequences
7	26	100.0	1063	6	BD273783 Sequences
8	26	100.0	1063	6	AX004916 Sequences
9	26	100.0	1243	6	BD273777 Sequences
10	26	100.0	1243	6	AX004882 Sequences
11	26	100.0	14664	1	AE0065931 Mycobacte
12	26	100.0	341957	15	EX842572 Mycobacte
13	26	100.0	343050	1	EX248334 Mycobacte
14	24	92.3	498	6	BD273782 Sequences
15	24	92.3	498	6	AX004914 Sequences
16	20.2	77.7	12021	1	AE005073 Halobacte
17	19.6	75.4	1497	8	AF124817 Mentha x
18	19.6	75.4	5906	1	AF522461 Mycobacte
19	19.6	75.4	16697	1	AE007137 Mycobacte
20	19.6	75.4	308050	1	EX248345 Mycobacte
21	19.6	75.4	349563	1	EX842582 Mycobacte
22	19.2	73.8	11791	3	AB035891 Drosophil
23	19.2	73.8	20604	1	AE008847 Salmonell
24	19.2	73.8	49445	2	AC015143 Drosophil
25	19.2	73.8	110000	3	AE003536_1 Continuation (2 of
26	19.2	73.8	173509	3	AC010031 Drosophil
27	19.2	73.8	216041	2	AC112524 Drosophil
28	18.8	72.3	1482	6	A98985 Sequence 3
29	18.8	72.3	3700	9	AF196313 Homo sapi
30	18.8	72.3	204767	2	AC011367 Homo sapi
31	18.8	72.3	254730	9	AC008533 Homo sapi
32	18.8	72.3	254730	9	AC008533 Homo sapi
33	18.6	71.5	129	6	AK101878 Sequence
34	18.6	71.5	4817	6	AX358525 Sequence
35	18.6	71.5	8933	6	AX358524 Sequence
36	18.6	71.5	20092	1	AE008804 Salmonell
37	18.6	71.5	140330	9	AC026777 Homo sapi
38	18.6	71.5	141370	9	AC010319 Homo sapi
39	18.6	71.5	154746	14	HSV2HG52 Z86099 Herpes simp
40	18.6	71.5	198344	2	AC087872 Mus muscu
41	18.6	71.5	256050	1	AL627274 Salmonell
42	18.6	71.5	300928	1	AE016836 Salmonell
43	18.2	70.0	176	8	HVU234864 Hordeum v
44	18.2	70.0	789	11	BV064586 S212P6390
45	18.2	70.0	110000	2	LMFLCHR32.10 Continuation (11 o

ALIGNMENTS

RESULT 1

BD273780

LOCUS

DEFINITION

Sequences nucleic acid from polypeptides exportes from

mycobacteries, vector the compranent and applications at diagnostic

and the prevention from the tuberculose.

BD273780

ACCESSION

BD273780.1 GI:33083548

VERSION

JP 2002534956-A/4.

KEYWORDS

Mycobacterium tuberculosis

SOURCE

Mycobacterium tuberculosis

ORGANISM

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium

tuberculosis complex.

REFERENCE

1 (bases 1 to 959)

Pred. No. is the number of results predicted by chance to have a

AUTHORS Gicquel,B., Portnoi,D., Lim,E., Pelicic,V., Guigueno,A. and
Salmoniere,Y.G.D.L.
TITLE Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the comprénant and applications at diagnostic
and the prevention from the tuberculose
JOURNAL Patent: JP 2002534956-A 4 22-OCT-2002;
INSTITUT PASTEUR
COMMENT OS Mycobacterium tuberculosis
PN JP 2002534956-A/4
PD 22-OCT-2002
PF 14-AUG-1998 JP 2000509849
PR 14-AUG-1997 FR 97/10404,11-SEP-1997 FR 97/11325 PI
BRIGITTE GICQUEL,DENIS PORTNOI,ENG-MONG LIM,VLADIMIR PELICIC, PI
AGNES GUIGUENO,YVES GOGUET DE LA SALMONIERE
PC C12N15/09,A61K39/04,A61K48/00,A61P11/00,A61P31/04,C07K14/35,
PC C07K16/12,
PC C07K19/00,C12N1/21,C12P21/02,C12Q1/02,C12Q1/68,G01N33/50, PC
G01N33/53//
PC (C12N15/09,C12R1.32),C12N15/00,(C12N15/00,C12R1.32) CC seq
ID no. 1C' Location/Qualifiers
FH Key (1)..(957).
FT CDS

FEATURES
source 1..959 Location/Qualifiers
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"

ORIGIN
Query Match 100.0%; Score 26; DB 6; Length 959;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATGACGGATCGTCGGGGCGGGAAC 26
|||||
Db 570 GATGACGGATCGTCGGGGCGGGAAC 595
|||||

RESULT 2
AX004906
LOCUS 959 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 25 from Patent W09909186.
ACCESSION AX004906
VERSION AX004906.1 GI:9928293
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE 1
AUTHORS Portnoi,D. and Guigueno,A.
TITLE Polypeptide nucleic sequences exported from mycobacteria, vectors
comprising same and uses for diagnosing and preventing tuberculosis
JOURNAL Patent: WO 9909186-A 25-25-FEB-1999;
PORTNOI DENIS (FR); GUIGUENO AGNES (FR)
FEATURES
source 1..959 Location/Qualifiers
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"
<1..201
/notes="unnamed protein product; seq ID no 1C"

CDS
/codon_start=1
/transl_table=11
/protein_id="CAC04929.1"
/db_xref="GI:9928293"
/db_xref="RENTREMBL:CAC04929"
/translation="SPTPIRSRSLATSLVPLRASSRSDTTYGRSGGGRHWVLGA
SVTPGSSLAICRSSATLSMSW"

ORIGIN
Query Match 100.0%; Score 26; DB 6; Length 959;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATGACGGATCGTCGGGGCGGGAAC 26
|||||
Db 570 GATGACGGATCGTCGGGGCGGGAAC 595
|||||

RESULT 3
BD273779
LOCUS 960 bp DNA linear PAT 17-JUL-2003
DEFINITION Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the comprénant and applications at diagnostic
and the prevention from the tuberculose.
ACCESSION BD273779.1 GI:33083547
VERSION BD273779.1
KEYWORDS JP 2002534956-A/3.
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE 1 (Bases 1 to 960)
AUTHORS Gicquel,B., Portnoi,D., Lim,E., Pelicic,V., Guigueno,A. and
Salmoniere,Y.G.D.L.
TITLE Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the comprénant and applications at diagnostic
and the prevention from the tuberculose
JOURNAL Patent: JP 2002534956-A 3 22-OCT-2002;
INSTITUT PASTEUR
COMMENT OS Mycobacterium tuberculosis
PN JP 2002534956-A/3
PD 22-OCT-2002
PF 14-AUG-1998 JP 2000509849
PR 14-AUG-1997 FR 97/10404,11-SEP-1997 FR 97/11325 PI
BRIGITTE GICQUEL,DENIS PORTNOI,ENG-MONG LIM,VLADIMIR PELICIC, PI
AGNES GUIGUENO,YVES GOGUET DE LA SALMONIERE
PC C12N15/09,A61K39/04,A61K48/00,A61P11/00,A61P31/04,C07K14/35,
PC C07K16/12,
PC C07K19/00,C12N1/21,C12P21/02,C12Q1/02,C12Q1/68,G01N33/50, PC
G01N33/53//
PC (C12N15/09,C12R1.32),C12N15/00,(C12N15/00,C12R1.32) CC seq
ID no. 1B' Location/Qualifiers
FH Key (1)..(960).
FT CDS

FEATURES
source 1..960 Location/Qualifiers
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"

ORIGIN
Query Match 100.0%; Score 26; DB 6; Length 960;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GATGACGGATCGTCGGGGCGGGAAC 26
|||||
Db 571 GATGACGGATCGTCGGGGCGGGAAC 596
|||||

RESULT 4
AX004895
LOCUS 960 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 14 from Patent W09909186.
ACCESSION AX004895
VERSION AX004895.1 GI:9928291
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

```

REFERENCE
AUTHORS      Portnoi,D. and Guigueno,A.
TITLE        Polypeptide nucleic sequences exported from mycobacteria, vectors
              comprising same and uses for diagnosing and preventing tuberculosis
JOURNAL      Patent: WO 9909186-A 14 25-FEB-1999;
              PORTNOI DENIS (FR); GUIGUENO AGNES (FR)
FEATURES
source       Location/Qualifiers
              1..960
              /organism="Mycobacterium tuberculosis"
              /mol_type="unassigned DNA"
              /db_xref="taxon:1773"
              <1..174
              /note="unnamed protein product; seq ID no 1B'"
              /codon_start=1
              /transl_table=11
              /protein_id="CAC04928.1"
              /db_xref="GI:9928292"
              /db_xref="REMBL:CAC04928"
              /translation="IAFDAYSVAQLFGDVTGARVAGVQPRHHIRPVLRLGPLGVGVC
              LHPRFAGYLSQ"
ORIGIN
Query Match      100.0%; Score 26; DB 6; Length 960;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATGACGGATCGTCGGGGCGGGAAC 26
Db 571 GATGACGGATCGTCGGGGCGGGAAC 596

RESULT 5
BD273778
LOCUS          BD273778          961 bp          DNA          linear          PAT 17-JUL-2003
DEFINITION    Sequences nucleic acid from polypeptides exportees from
              mycobacteries, vector the complement and applications at diagnostic
              and the prevention from the tuberculose.
ACCESSION     BD273778.1 GI:33083546
VERSION       JP 2002534956-A/2
KEYWORDS      Mycobacterium tuberculosis
SOURCE        Mycobacterium tuberculosis
ORGANISM      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE     1 (bases 1 to 961)
AUTHORS      Gicquel,B., Portnoi,D., Lim,E., Pelicic,V., Guigueno,A. and
              Salmoniere,Y.G.D.L.
TITLE        Sequences nucleic acid from polypeptides exportees from
              mycobacteries, vector the complement and applications at diagnostic
              and the prevention from the tuberculose
JOURNAL      Patent: JP 2002534956-A 2 22-OCT-2002;
              INSTITUT PASTEUR
COMMENT       OS Mycobacterium tuberculosis
              PN JP 2002534956-A/2
              PD 22-OCT-2002
              PP 14-AUG-1998 JP 2000509849
              PR 14-AUG-1997 FR 97/10404,11-SEP-1997 FR 97/11325 PI
              BRIGITTE GICQUEL,DENIS PORTNOI,ENG-MONG LIM,VLADIMIR PELICIC, PI
              AGNES GUIGUENO,YVES GOGUET DE LA SALMONIERE
              PC C12N15/09,A61K39/04,A61K48/00,A61P11/00,A61P31/04,C07K14/35,
              PC C07K16/12,
              PC C07K19/00,C12N1/21,C12P21/02,C12Q1/02,C12Q1/68,G01N33/50, PC
              G01N33/53//
              PC (C12N15/09,C12R1:32),C12N15/00,(C12N15/00,C12R1:32) CC seq
              id no. 1A'
FH key        Location/Qualifiers
FT CDS        (1)..(960).
              1..961
              /organism="Mycobacterium tuberculosis"
              /mol_type="genomic DNA"
              /db_xref="taxon:1773"

```

```

ORIGIN
Query Match      100.0%; Score 26; DB 6; Length 961;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATGACGGATCGTCGGGGCGGGAAC 26
Db 572 GATGACGGATCGTCGGGGCGGGAAC 597

RESULT 6
BD273783
LOCUS          BD273783          1063 bp          DNA          linear          PAT 17-JUL-2003
DEFINITION    Sequences nucleic acid from polypeptides exportees from
              mycobacteries, vector the complement and applications at diagnostic
              and the prevention from the tuberculose.
ACCESSION     BD273783.1 GI:33083551
VERSION       JP 2002534956-A/7.
KEYWORDS      Mycobacterium tuberculosis
SOURCE        Mycobacterium tuberculosis
ORGANISM      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE     1 (bases 1 to 1063)
AUTHORS      Gicquel,B., Portnoi,D., Lim,E., Pelicic,V., Guigueno,A. and
              Salmoniere,Y.G.D.L.
TITLE        Sequences nucleic acid from polypeptides exportees from
              mycobacteries, vector the complement and applications at diagnostic
              and the prevention from the tuberculose
JOURNAL      Patent: WO 9909186-A 8 25-FEB-1999;
              PORTNOI DENIS (FR); GUIGUENO AGNES (FR)
FEATURES
source       Location/Qualifiers
              1..961
              /organism="Mycobacterium tuberculosis"
              /mol_type="unassigned DNA"
              /db_xref="taxon:1773"
              <13..>197
              /note="unnamed protein product; seq id no 1A'"
              /codon_start=1
              /transl_table=11
              /protein_id="CAC04927.1"
              /db_xref="GI:9928290"
              /db_xref="REMBL:CAC04927"
              /translation="RLFGRAAFWRHWCRCRPAATPHTAGPVAGVAGWVFPSP
              QAVRWLFVAVARRHCRCL"
ORIGIN
Query Match      100.0%; Score 26; DB 6; Length 961;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATGACGGATCGTCGGGGCGGGAAC 26
Db 572 GATGACGGATCGTCGGGGCGGGAAC 597

RESULT 7
BD273783
LOCUS          BD273783          1063 bp          DNA          linear          PAT 17-JUL-2003
DEFINITION    Sequences nucleic acid from polypeptides exportees from
              mycobacteries, vector the complement and applications at diagnostic
              and the prevention from the tuberculose.
ACCESSION     BD273783.1 GI:33083551
VERSION       JP 2002534956-A/7.
KEYWORDS      Mycobacterium tuberculosis
SOURCE        Mycobacterium tuberculosis
ORGANISM      Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
              Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
              tuberculosis complex.
REFERENCE     1 (bases 1 to 1063)
AUTHORS      Gicquel,B., Portnoi,D., Lim,E., Pelicic,V., Guigueno,A. and
              Salmoniere,Y.G.D.L.
TITLE        Sequences nucleic acid from polypeptides exportees from
              mycobacteries, vector the complement and applications at diagnostic

```

```

and the prevention from the tuberculo
Patent: JP 2002534956-A 7 22-OCT-2002;
INSTITUT PASTEUR
OS Mycobacterium tuberculosis
PN JP 2002534956-A/7
PD 22-OCT-2002
PF 14-AUG-1998 JP 2000509849
PR 14-AUG-1997 FR 97/10404,11-SEP-1997 FR 97/11325 PI
BRIGITTE GICQUEL,DENIS PORTNOI,ENG-MONG LIM,VLADIMIR PELICIC, PI
AGNES GUIGUENO,YVES GOGUET DE LA SALMONIERE
PC C12N15/09,A61K39/04,A61K48/00,A61P11/00,A61P31/04,C07K14/35,
C07K16/12
PC C07K19/00,C12N1/21,C12P21/02,C12Q1/02,C12Q1/68,G01N33/50, PC
GOIN33/53//
PC (C12N15/09,C12R1:32),C12N15/00,(C12N15/00,C12R1:32) CC seq
ID no. 2
FH Key Location/Qualifiers
FT CDS (53)..(82)
FT CDS (180)..(437)
FT CDS (451)..(858)
FT CDS (862)..(1062).
FEATURES
source
1..1063
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"
ORIGIN
Query Match 100.0%; Score 26; DB 6; Length 1063;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATGACGGATCGTCGGGGGGGGAAC 26
|||||
DB 362 GATGACGGATCGTCGGGGGGGGAAC 387
RESULT 8
AX004916 1063 bp DNA linear PAT 24-AUG-2000
LOCUS
DEFINITION
Sequence 35 from Patent WO9909186.
ACCESSION
AX004916
VERSION
AX004916.1 GI:9928299
SOURCE
Mycobacterium tuberculosis
ORGANISM
Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1
REFERENCE
AUTHORS
Portnoi,D. and Guigueno,A.
TITLE
Polypeptide nucleic sequences exported from mycobacteria, vectors
comprising same and uses for diagnosing and preventing tuberculosis
Patent: WO.9903186-A 35 25-FEB-1999;
JOURNAL
PORTNOI DENIS (FR); GUIGUENO AGNES (FR)
FEATURES
source
1..1063
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"
53..85
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC04932.1"
/db_xref="GI:9928300"
/db_xref="REMTREMBL:CAC04932"
/translation="MEVELVDDQ"
180..440
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC04933.1"
and the prevention from the tuberculo
Patent: JP 2002534956-A 7 22-OCT-2002;
INSTITUT PASTEUR
OS Mycobacterium tuberculosis
PN JP 2002534956-A/7
PD 22-OCT-2002
PF 14-AUG-1998 JP 2000509849
PR 14-AUG-1997 FR 97/10404,11-SEP-1997 FR 97/11325 PI
BRIGITTE GICQUEL,DENIS PORTNOI,ENG-MONG LIM,VLADIMIR PELICIC, PI
AGNES GUIGUENO,YVES GOGUET DE LA SALMONIERE
PC C12N15/09,A61K39/04,A61K48/00,A61P11/00,A61P31/04,C07K14/35,
C07K16/12
PC C07K19/00,C12N1/21,C12P21/02,C12Q1/02,C12Q1/68,G01N33/50, PC
GOIN33/53//
PC (C12N15/09,C12R1:32),C12N15/00,(C12N15/00,C12R1:32) CC seq
ID no. 2
FH Key Location/Qualifiers
FT CDS (53)..(82)
FT CDS (180)..(437)
FT CDS (451)..(858)
FT CDS (862)..(1062).
FEATURES
source
1..1063
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"
ORIGIN
Query Match 100.0%; Score 26; DB 6; Length 1063;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATGACGGATCGTCGGGGGGGGAAC 26
|||||
DB 362 GATGACGGATCGTCGGGGGGGGAAC 387
RESULT 9
BD273777 1243 bp DNA linear PAT 17-JUL-2003
LOCUS
DEFINITION
Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the compranat and applications at diagnostic
and the prevention from the tuberculose.
BD273777
ACCESSION
BD273777.1 GI:33083545
VERSION
JP 2002534956-A/1.
KEYWORDS
Mycobacterium tuberculosis
SOURCE
Mycobacterium tuberculosis
ORGANISM
Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1 (bases 1 to 1243)
REFERENCE
AUTHORS
Gicquel,B., Portnoi,D., Lim,E., Pelicic,V., Guigueno,A. and
Salmoniere,Y.G.D.L.
TITLE
Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the compranat and applications at diagnostic
and the prevention from the tuberculose
JOURNAL
Patent: JP 2002534956-A 1 22-OCT-2002;
INSTITUT PASTEUR
OS Mycobacterium tuberculosis
PN JP 2002534956-A/1
PD 22-OCT-2002
PF 14-AUG-1998 JP 2000509849
PR 14-AUG-1997 FR 97/10404,11-SEP-1997 FR 97/11325 PI
BRIGITTE GICQUEL,DENIS PORTNOI,ENG-MONG LIM,VLADIMIR PELICIC, PI
AGNES GUIGUENO,YVES GOGUET DE LA SALMONIERE
PC C12N15/09,A61K39/04,A61K48/00,A61P11/00,A61P31/04,C07K14/35,
C07K16/12,
PC C07K19/00,C12N1/21,C12P21/02,C12Q1/02,C12Q1/68,G01N33/50, PC
GOIN33/53//
PC (C12N15/09,C12R1:32),C12N15/00,(C12N15/00,C12R1:32) CC seq
ID no. 1
FH Key Location/Qualifiers
FT CDS (21)..(38)
FT CDS (124)..(453)
FT CDS (473)..(481)

```



```

FT CDS (573) . . (671)
FT CDS (719) . . (928)
FT CDS (942) . . (1241)
FEATURES
    source
        Location/Qualifiers
            1. .1243
                /organism="Mycobacterium tuberculosis"
                /mol_type="genomic DNA"
                /db_xref="taxon:1773"
ORIGIN
    Query Match 100.0%; Score 26; DB 6; Length 1243;
    Best Local Similarity 100.0%; Pred. No. 2.4;
    Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATGACGGATCGTCGGGGCGGGGAAC 26
    |||||
DB 853 GATGACGGATCGTCGGGGCGGGGAAC 878
    |||||
RESULT 10
LOCUS AX004882 1243 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 1 from Patent WO9909186.
ACCESSION AX004882
VERSION AX004882.1 GI:9928282
KEYWORDS
SOURCE
    Mycobacterium tuberculosis
    Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
    Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
    tuberculosis complex.
REFERENCE
    1
    Portnoi, D. and Guigueno, A.
    Polypeptide nucleic sequences exported from mycobacteria, vectors
    comprising same and uses for diagnosing and preventing tuberculosis
    Patent: WO 9909186-A 1 25-FEB-1999;
    PORTNOI DENIS (FR); GUIGUENO AGNES (FR)
FEATURES
    source
        Location/Qualifiers
            1. .1243
                /organism="Mycobacterium tuberculosis"
                /mol_type="unassigned DNA"
                /db_xref="taxon:1773"
                21. .41
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC04921.1"
                /db_xref="GI:9928283"
                /db_xref="REMTREMBL:CAC04921"
                /translation="MVVGMT"
                124. .456
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC04922.1"
                /db_xref="REMTREMBL:CAC04922"
                /db_xref="REMTREMBL:CAC04922"
                /translation="WAGDOELRFDVPLYLTAESRYLVVPRATLWADGYERRPA
                NAPVQGPQIAFDAYSVAQLFGDVTGARVAGVQQRHHPVRLRGLGVCGLRHPR
                QPAGVLSQ"
                473. .484
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC04923.1"
                /db_xref="GI:9928285"
                /db_xref="REMTREMBL:CAC04923"
                /translation="MSW"
                573. .674
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC04924.1"
                /db_xref="GI:9928286"
                /db_xref="REMTREMBL:CAC04924"
                /translation="MTATRLNRHRLDPTASSPGKPPALTATNP"
                719. .931
                /note="unnamed protein product"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC04925.1"
                /db_xref="GI:9928287"
                /db_xref="REMTREMBL:CAC04925"
                /translation="MNCWISDPSVRAVRAREPTDRVHAFVGDRTAPGVGABRGD
                RMTDRRGRLPGRRTVANPSQTRKP"
                942. .1241
                /note="unnamed protein product; seq ID nol"
                /codon_start=1
                /transl_table=11
                /protein_id="CAC04926.1"
                /db_xref="GI:9928288"
                /db_xref="REMTREMBL:CAC04926"
                /translation="MKTGTATRRLLAVLLALPGAAVALLAEPATGASDPCAAS
                EVATGVSVAKSMGDIYLDSPETNQWTAVLQQVQGGVSVASLKAHFAEPKVASD"
ORIGIN
    Query Match 100.0%; Score 26; DB 6; Length 1243;
    Best Local Similarity 100.0%; Pred. No. 2.4;
    Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GATGACGGATCGTCGGGGCGGGGAAC 26
    |||||
DB 853 GATGACGGATCGTCGGGGCGGGGAAC 878
    |||||
RESULT 11
LOCUS AE006931 14664 bp DNA linear BCT 27-APR-2001
DEFINITION Mycobacterium tuberculosis CDC1551, section 17 of 280 of the
    complete genome.
ACCESSION AE006931 AE000516
VERSION AE006931.1 GI:13879697
KEYWORDS
SOURCE
    Mycobacterium tuberculosis CDC1551
    Mycobacterium tuberculosis CDC1551
    Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
    Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
    tuberculosis complex.
REFERENCE
    1 (bases 1 to 14664)
    Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
    Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
    Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
    Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
    Gill, J., Mikula, A. and Bishai, W.
    Whole genome comparison of Mycobacterium tuberculosis clinical and
    laboratory strains
    Unpublished
JOURNAL
REFERENCE
    2 (bases 1 to 14664)
    Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
    Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
    Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
    Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
    Gill, J., Mikula, A. and Bishai, W.
    Direct Submission
    Submitted (25-APR-2001) The Institute for Genomic Research, 9712
    Medical Center Dr, Rockville, MD 20850, USA
    Location/Qualifiers
        1. .14664
            /organism="Mycobacterium tuberculosis CDC1551"
            /mol_type="genomic DNA"
            /strain="CDC1551"
            /db_xref="taxon:83331"
            /note="clinical strain"
            142. .552
            /Gene="WT0213"
            142. .552
            /Gene="WT0213"
FEATURES
    source
        /organism="Mycobacterium tuberculosis CDC1551"
        /mol_type="genomic DNA"
        /strain="CDC1551"
        /db_xref="taxon:83331"
        /note="clinical strain"
        142. .552
        /Gene="WT0213"
        142. .552
        /Gene="WT0213"
gene
    /Gene="WT0213"
CDS
    142. .552
    /Gene="WT0213"

```

```

/gene="MT0213"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK4434.1"
/db_xref="GI:13879598"
/translation="MTGTATTTRRLRLAVLIALPAAVALLAAPSATGSDPPCAAS
EAVTGVSAKSGDYLDSHPETNQVMTAVLQQVPGPGSVASLKAHFAFPKVASDLH
ALSQPLTDLSTCRSLPSISGLQALQMAVQGARR"
/complement(604..1842)
/gene="MT0214"
/complement(604..1842)
/gene="MT0214"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK4435.1"
/db_xref="GI:13879599"
/translation="MSHDAPARNLRQVRGALPRTRVAPPAGVPVPRGKYWLRWAVL
AIVAIIVALEVALGMDQAKAVSYRAKWMWLLAAVAAAGASMSHFAIQRTLLKSA
GHVVKQWRSEAFYAANSLSSTLPGGPVLSTFLLRQRIWGASTVVASQVLMVSGVL
QAVGLALLGLGGAFFPKAKNPFLFTLGGFTLLLAQVASTPELTIGIGRRVLS
WANSVRGPADAGLKWRRTLMQLESVSLGRDLGVAFGMSLFNMIADVACLGFPAAY
AGHASVGGIAYAYARAVGTIPLMPGGLLYVEAVLVECLVSSCMPLSAISMLIY
RLISWLLIIAAGVNVFFFRFTSTADSNDRDPFDENLRUVIQGTGCPDDPVETT
PQGPAPTPDLRPGGTPPR"
/2012..3115
/gene="MT0215"
/2012..3115
/gene="MT0215"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK4436.1"
/db_xref="GI:13879700"
/translation="WSASLDDASVAPLVEKTAARWRLAVILAAVALLVWLKPEVI
VVPVLLALMSALLVPPVDPDLDSRGLPHAVATVTLVLSGFVGLGILTFVVSQFIAG
PHLTVVERSIDSARWLEIEGPAHURGEQIDNAGNAEALRNKQKLTGSLSTPAAT
ITELVATVAVLVFTLFIYLGGRSQQWYTKAFPSVDRVRAAGRAGYASLIGYARA
TELVALTDAAGVAGLVNGLFALPLASLFFGAFIPILGVAVAGFLVAVVALLAG
IGVATLIGVLLAVNLEAHLLQPLVMGRVSIHPLAVVLAIAAGGVLAGVVGALLAV
PTVAFNNVAVOVLGNGNPDADVSSDLTEV"
/complement(3112..5946)
/gene="MT0216"
/complement(3112..5946)
/gene="MT0216"
/notes="similar to GP:2104606; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, MmpL family"
/protein_id="AAK4437.1"
/db_xref="GI:13879701"
/translation="MPAWGRTVRYRFTVIGVMVALCLGGVGLSGHVTQSGFY
DQGSQVQASVIGDQVYGRDTSCHIVAIFQAPAGKTVDVDPMSKKVVDENARFQDHP
DOVLGWAGVLRASQATGATADCKTFFVSIPLKGDGDDDTILNNYKAIAPDLQRLDGT
VKLAGIQVPAEALTGTIATDQRMELVIALPVAVLFFVPGGVIAAGLPVWVGVCIA
GALGNRPLIAGPVHYPAQPVVSLIGLGIADYGLFTVSRPREIASGYDETAVRR
VTITAGRTVTFSAVILVAGAILLFPQGLKSLTATIASVMSLSILSITVLPACLG
ILGHVDAVGRTLFVFLPLANKKISAAYLNWDLRLQRTREEVEAGFVWGLKNRV
MRKPLVAPVIMILLIIPVGLSLGISEKIPPTNSVRQAQEEFDKLPFGYRTN
PLVLTQTNHNPQVTDQADIRSKMAIGGIEPDNDPANMQERAVAGASQDPS
RVLQNLINPADASKLLEALTPPKGTIVLVGDPDALELDSIHGLFAKPLMVVIL
LTTIVLFLAGSVLPKATLMSALTGSTWGLITWIFVDGHSKWLNTPTPLTA
PHIGLIIALVFLGSTDYEVFLSVRERKMSQEBAIRITANTAGLIIIAALIIYA
VVAGAFVSDLVMMKTLAFGLMAALLDDATVFMFLVPSVVKLLGDDCWVWARRL
QTRIGELHLPEDRNKXVNGRPPVPTAGLVNARAGDPPPHDPHPLSPRP
ARSSPASSPELTAAAPASGASTTRMQIGSTTEPTPTRLAARQVQSPAST
PPPTPTPPGAFSAGQTRAMPANRSTDAAGDPABPTAALPTIIRSDGDDSEATQNLN
ARGTSDTKTRRRGGGALSQDOLLREGRL"

```

```

complement(6012..6746)
/gene="MT0217"
complement(6012..6746)
/gene="MT0217"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK4438.1"
/db_xref="GI:13879702"
/translation="MMSLTDEVTSTSELSARHSLAEDLSODGLTSLCAPGARVLL
VMDPNLDMGLSILGRPTALRPRFDALGRWLLARTARIVAGRGTSTPEATVFTT
NIAPGSAEVVRPWVLDALRNVGFAVFAKPKVDESDVDROMLAHIDRERYEGLAALVA
SACQAPRFLAEVARSQVQVIGFRESHASWALSOTLEFVDELDIAGVFREPLPRI
GLDSLPEQAWLQPFRRPLSLSLTSRV"
/complement(6743..7534)
/gene="MT0218"
/complement(6743..7534)
/gene="MT0218"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK4439.1"
/db_xref="GI:13879703"
/translation="MVHGGOMHAQPGVGLRPTPVASGQLPSTSISSRSIGSKAORE
TWERLWPELGLLALPQSPRGTPVDTRAWFGRDAPVLEIGSGSGTSTLAMAKAEPRVD
VIADVVRHKLQALLCAIDKVGSDGINRILGNADVVDVLOHLIAPDSLCGVRVFFDDP
WPKARHKKRLQPATMALIADRLVPSGVLHAATDHPYAEHIAAAGDAEPRLVRVDP
DELLPISVVRPATKYERKAQGGAVIELLWKKGCSERDLKIR"
/7666..8751
/gene="MT0219"
/7666..8751
/gene="MT0219"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK4440.1"
/db_xref="GI:13879704"
/translation="MRGQHQIIVDELARPATSSAQORVVAIAQRAAEPLRVAVGRPP
GVGTRVARALQAGSSGGMVTVPQARAADSDVLYVYVTVVEVVKPEDEAIAATRRP
WVAVLNKADLAPLGSAGPTVMAQACQAFSTLLGVPMESMIGLVALLAALDDDDTLR
AVRRLAAAPDGFDAIDRAVAGLAAALPVTEVRLLRLDLDLFLGALGMARFPRGR
PSTPAQRLTLRRVSGDAVIDKVTAAAGSEVRYRLLDAVAELEALAAQAKEIGGPI
GEFLRDDTVLARMAAADVVALVGLDVGPLDDPAAHLPRAVRWHRYSLDNGDMHRTC
GADIARGSLSLNSLAGMPLHYRKSS"
/8781..10226
/gene="MT0220"
/8781..10226
/gene="MT0220"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK4441.1"
/db_xref="GI:13879705"
/translation="MDELVAIAAPFAGLAGLGPVIRREVVLTGTPWLAGVSGVRAALA
ERLPQRRFVTAELGPGDAPVAVVSVSAATALTESDCLVLLDTAAEHTDAVAVVSKI
DVHRGMDVLTNSRDLARASRYARVPMVAAAAPLGEPLYDLDLVAAIQKOLADPA
VARNRMWESRLLMVARFDCDAQSAGRARVDALROQRTVLRQGRQSKSEHTIA
LRAQIQHARVKSIFYARNRCSLLRVLQEHVAGLSKDIARFAAYTRGRVQEVVAEVG
EGAVAHADVAGLQVGVQVQPVLENIPAVLFTVVPAPLTSRRLEITRLLGLAGFLG
IAUTLRVAGLTPGLAAGMAGVAGIATVAVVNAARALLHDRVVRVDRWTEGVTAS
LRSVQELATRVVAVETLSTAISSRDDAENARVADQVSIIDGELLEHAFAAARAA
LRDREMPVAPAALEAVRAELGEPGATTLGF"
/10410..12230
/gene="MT0221"
/10410..12230
/gene="MT0221"
/notes="similar to PID:2104609; identified by sequence
similarity; putative"
/codon_start=1

```

```
/transl table=11
/product="phosphoenolpyruvate carboxykinase"
/protein_id="AAK4442.1"
/db_xref="GI:13879706"
/translation="MTSATPGLDTPATNHQGLLSWVEVAELTQDPRVYVFDGSEEE
FORLQDLVEAGTFFIRLNPEKKNYSYALSDPSVARVESRTYICSAKIDAGPTNNW
MDPGENRSIMKDIYRCMRGRMTYVVFQWGLGADPKLGVETIDSEYVVVSMRTMT
RMGALEKMGDDGPFVKALHSVGAFLPQGXDAWPCSETKYTHFFPBTIRIWSYGS
GYGNALIGKCKYSYLAISAMAHDEGLAEHMLIKLJSPENKAYVFAAFSPACGKT
NLMLQPTIPGRAETLIGDDIAMRFGKGRGLYANPFGFVGAPGCTNWKSNPNMR
TIAAGNTFTNVALTDGDDVWMEGLSDFOHLIDMGNDWYFRETETNAAHNSRYCT
PMSQCPIIAPENDPQGVPTISGLFGRRKTVPLVTEARDWQHGFIGATIGSSQTA
AAEGKVGNRVDRPMALPFLGVNGDYFORWNLGHADESKLPKVFVFNWFRGSDG
RFLWPGFGENSRLKWIVDRIEHKAGCATTPGTVPADVLDLGLDVAADVAALA
VDADENRQELPLIEEMLQFVGEKLPFGVXDEFALKERLG"

Query Match      100.0%; Score 26; DB 1; Length 14564;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGAAC 26
DB 53 GATGACGGATCGTCGGGGCGGGAAC 78

|||||
|||||

RESULT 12
BX842572 standard; circular genomic DNA; PRO; 341957 BP.
AC BX842572; AL021427; AL021428; AL021926; AL021927; AL021928; AL021929;
AC AL021930; AL124456; Z74410; Z80233; Z80775; Z85089; Z92669; Z92770; Z96071;
AC Z97050;
SV BX842572.1
DT 21-NOV-2003 (Rel. 77, Created)
DT 21-NOV-2003 (Rel. 77, Last updated, Version 1)
XX Mycobacterium tuberculosis H37Rv complete genome; segment 1/13
XX complete genome.
KW
OS Mycobacterium tuberculosis H37Rv
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium;
XX Mycobacterium tuberculosis complex; Mycobacterium tuberculosis.
[1]
RX MEDLINE; 98295987.
RX PUBMED; 9634230.
RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry III C.E., Tekala F., Badcock K.,
RA Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K.,
RA Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K.,
RA Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J.,
RA Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S.,
RA Squares S., Squires R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete
RT genome sequence";
RL Nature 393:537-544 (1998).
[2]
RX PUBMED; 12368430.
RA Camus J.C., Pryor M.J., Medigue C., Cole S.T.;
RT "Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv";
RL Microbiology 148:2967-2973 (2002).
[3]
RX 1-341957
RA Parkhill J.;
RT ;
RL Submitted (11-JUN-1998) to the EMBL/GenBank/DBJ databases.
RL Submitted on behalf of the Mycobacterium tuberculosis sequencing and
```

```
RL mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton,
RL Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut
RL Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail:
RL parkhill@sanger.ac.uk
XX
CC Details of M. tuberculosis sequencing at the Sanger Centre
CC are available on the World Wide Web.
CC (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/)
XX
Key Location/Qualifiers
source 1..341957
/db_xref="taxon:83332"
/mol_type="genomic DNA"
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
1..1524
/evidence=EXPERIMENTAL
/note="RV0001, (MT0001, MT0029.01, P49993), len: 507 aa.
dnaA, chromosomal replication initiator protein (see
citations below), equivalent to other Mycobacterial
CHROMOSOMAL REPLICATION INITIATOR PROTEINS e.g.
P46388|DNAA_MYCLE from Mycobacterium leprae (502 aa);
Q9L7L7|DNAA_MYCPA from Mycobacterium paratuberculosis (509
aa); P49990|DNAA_MYCAV from Mycobacterium avium (508 aa);
P49992|DNAA_MYCSM from Mycobacterium smegmatis (504 aa);
etc. Also highly similar to others except in N-terminus
e.g. O9ZH75|DNAA_STRCH CHROMOSOMAL REPLICATION INITIATOR
PROTEIN from Streptomyces chrysomallus (624 aa);
Q9ZH76|DNAA_STRRE from Streptomyces reticuli (643 aa);
DNAA_ECOLI|P03004|B3702 chromosomal replication initiator
protein from Escherichia coli strain K12 (467 aa), FASTA
scores: opt: 986, E(): 0, (43.2% identity in 389 aa
overlap); etc. Contains P500017 ATP/GTP-binding site motif
A (P-loop) and P501008 DnaA protein signature. BELONGS TO
THE DNAA FAMILY. Note that the first base of this gene has
been taken as base 1 of the Mycobacterium tuberculosis
H37Rv genomic sequence."
/transl_table=11
/function="PLAYS AN IMPORTANT ROLE IN THE INITIATION AND
REGULATION OF CHROMOSOMAL REPLICATION. BINDS TO THE ORIGIN
OF REPLICATION; IT BINDS SPECIFICALLY DOUBLE-STRANDED DNA
AT A 9 BP CONSENSUS (DNAA BOX): 5'-TTATC(C/A)(C/A)A-3'.
DNAA BINDS TO ATP AND TO ACIDIC PHOSPHOLIPIDS. DNAA PROTEIN
BINDS THE ORIGIN OF REPLICATION (oriC), ATP AND ADP, AND
EXHIBITS WEAK ATPase ACTIVITY."
/gene="dnaA"
/locus_tag="RV0001"
/product="CHROMOSOMAL REPLICATION INITIATOR PROTEIN DNAA"
/translation="MTDDPGSGFTVNVAVSELNGDPKVDGSPSSDANLSAPLTPQOR
AWMLVQPLTIVEGALLSPVSFVQNEIERHRAPIIDALSRELHQIQGVRIAPPA
TDEADDTTPPSNPATTSPTTDDNDEIDDSAAAGDNCNHSWPSYFTEPHNTDSATA
GYSLNRYTFDTFVIGASNRFAHAALAAEAPAYANPLFIWGESLGTGHTLHAAG
NVTSLRPGMKVYVSTEFNTDINSRDORRKAQKRSYVDVLLVDDIQTLEGKEG
ICSEFEFTNTHANKOVLISDPBPQOLATLEDRLTREMGLITDTPQPELETRA
ILRKAQMERLAVFDLVLELIASSIERNIRELEGALIRVTAFAINKTPIDKALAEIVL
ILKADANTMQISAATNAATAYFTDTEBELRGFKTRALAQSRQIATMYLCRELTDLS
LPKICQAFGRDHTTVMYAQRKILSEMAERREVFDFVXELTIRIQSKR"
622..645
/note="P500017 ATP/GTP-binding site motif A (P-loop)"
1384..1440
/note="P501008 DnaA protein signature"
2052..3260
/evidence=EXPERIMENTAL
/note="RV0002, (MTV029.02, MTCY10H4.0), len: 402 aa. dnaN,
DNA polymerase III (beta chain) (EC 2.7.7) (see citations
below), equivalent to other Mycobacterial DNA POLYMERASES
III BETA CHAIN e.g. NP_301130.1|NC_002677 from
Mycobacterium leprae (399 aa); Q9L7L6|DP3B_MYCPA from
Mycobacterium avium subsp. paratuberculosis (399 aa);
P52851|DP3B_MYCSM from Mycobacterium smegmatis (397 aa);
```

etc. Also highly similar to others e.g. P27903|DP3B_STRCO
DNA POLYMERASE III BETA CHAIN from Streptomyces coelicolor
(376 aa), FASTA scores: Opt: 1189, E(): 0, (52.8% identity
in 337 aa overlap); P21174|DP3B_MICLU from Micrococcus
luteus (310 aa); P52023|DP3B_SYNPF from Synecococcus sp.
strain PCC 7942 (375 aa); etc. Overlaps and extends CDS in
neighbouring cosmid MTCY10H4.01."
/transl table=11
/EC_number="2.7.7.7"
/function="DNA POLYMERASE III IS A COMPLEX, MULTICHAIN
ENZYME RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN
BACTERIA. THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5'
EXONUCLEASE ACTIVITY. THE BETA CHAIN IS REQUIRED FOR
INITIATION OF REPLICATION ONCE IT IS CLAMPED ONTO DNA. IT
SLIDES FREELY (BIDIRECTIONAL AND ATP-INDEPENDENT) ALONG
DUPEX DNA [CATALYTIC ACTIVITY: N deoxynucleoside
triphosphate = N diphosphate + {DNA}N]."
/genes="dnaA"
/locus_tag="Rv0002"
/product="DNA POLYMERASE III (BETA CHAIN) DNAN (DNA
NUCLEOTIDYLTRANSFERASE) 1"
/protein_id="CAAL6239.1"
/translation="MDAATRVGLTDLTRELRESFADAVSWAKLPARPAPVPLSGV
LRTGSLGRLISGDFEVSAGAEIVPSGVSIVSGRLSDITRALFNKPDVHVEG
NRVALTCGNARSLPTMVEDYPTLPTPEETGLLPALFAEAIISOVAIAQRDDTPEM
LTGIRVEILGETVLAATDFRLAVRELKWSASSPIEAAVLVPKATLAFAAKAGIGSS
DVRSLGTGPGVKDGLGILGISGKSTTRELDAEPFKPQLLPTEHTAVATMDVAELI
EAIKVALVADRAQVVRMFADGVSRLSAGADVGAEDLVVDVAGEPLTIAFNPTYL
TDGLSSRSRVSGFTTAGKPALLPVSGDDRPVAGLNGGPFPAVSTIDVYLLMPVR
LPG"
3280..4437
/evidence=EXPERIMENTAL
/notes="Rv0003, (MTCY10H4.01), len: 385 aa. recF, DNA
replication and repair protein (see citations below),
equivalent to others Mycobacterial DNA replication and
repair proteins e.g. NP 301131.1|NC 002677 from
Mycobacterium leprae (385 aa); G9L7L5|REC_F_MCPA from
Mycobacterium avium subsp. paratuberculosis (385 aa);
P50916|REC_MYCSM from Mycobacterium smegmatis (384 aa);
etc. Also highly similar to others e.g. P36176|REC_F_STRCO
DNA REPLICATION AND REPAIR PROTEIN from Streptomyces
coelicolor (373 aa); NP 440892.1|NC 000911 from
Synecococcus sp. strain PCC 6803 (384 aa);
NP 469352.1|NC 003212 from Listeria innocua (370 aa); etc.
Contains P500017 ATP/GTP-binding site motif A (P-loop),
P500617 RecF protein signature 1, and P500618 RecF protein
signature 2. BELONGS TO THE REC_F FAMILY."
/transl table=11
/function="THE REC_F PROTEIN IS INVOLVED IN DNA METABOLISM
AND RECOMBINATION; IT IS REQUIRED FOR DNA REPLICATION AND
NORMAL SOS INDUCIBILITY. REC_F BINDS PREFERENTIALLY TO
SINGLE-STRANDED, LINEAR DNA. IT ALSO SEEMS TO BIND ATP."
/genes="recF"
/locus_tag="Rv0003"
/product="DNA REPLICATION AND REPAIR PROTEIN REC_F
(SINGLE-STRAND DNA BINDING PROTEIN)"
/protein_id="CA802424.1"
/translation="MYRHGLRDFRSWACVLELHPGRVTVFGNGYKTNLIALWY
STTLGHRVSADPLIRVGTDRAVISTIIVNDGRECAVDLEIATGRVKARLNRSVRS
TRDVGVLRVADLGLDVRGDPDRRYLLDDLAIVRPAIAVRAEYRVLQRQTA
LKSVPGRYGRGVFTDVLWDSRLAEGAEIVAAIRDLVNLQALPEVKYAYLLAPE
SRASIGYRSMVDTPGESDIDROLLAARLIALAARDAELRGVCLVGFPHRDDLI
LRIGDPAKFSHGSAWLSAVALRAAYQLLRVGGEPVLLDDVFAELDVVRRRALA
TAESAQVLVTHAVLEIDIPAGWDARRVIDVRADDTGSMSVLP"
3367..3390
/notes="P500017 ATP/GTP-binding site motif A"
3634..3690
/notes="P500617 RecF protein signature 1"
4243..4296
/notes="P500618 RecF protein signature 2"
4434..4497
/evidence=EXPERIMENTAL
/notes="Rv0004, (MTCY10H4.02), len: 187 aa. Conserved

hypothetical protein (see Salazar et al., 1996), highly
similar, but longer 21 aa in N-terminus, to
AAF33656.1|AF222789 unknown protein from Mycobacterium
avium subsp. paratuberculosis (166 aa); and highly similar
to NP_301132.1|NC 002677 conserved hypothetical protein
from Mycobacterium leprae (189 aa); S70890 hypothetical
protein from Mycobacterium smegmatis (194 aa). Also highly
similar, except in N-terminal part, to
Query Match 100.0%; Score 26; DB 15; Length 341957;
Best Local Similarity 100.0%; Pred. NO. 1.2; Indels 0; Gaps 0;
Matches 26; Conservative 0; Mismatches 0
QY 1 GATGACGATCTCGGGGGCGGGAAC 26
|||||
Db 241425 GATGACGATCTCGGGGGCGGGAAC 241450
|||||
RESULT 13
BX248334 343050 bp DNA linear BCT 11-JUN-2003
LOCUS Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment
1/14
DEFINITION BX248334 BX248333
ACCESSION BX248334 BX248333
VERSION BX248334.1 GI:31616762
KEYWORDS complete genome.
SOURCE Mycobacterium bovis subsp. bovis AF2122/97
ORGANISM Mycobacterium bovis subsp. bovis AF2122/97
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1
REFERENCE
AUTHORS
Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H.,
Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsemp, C., Simon, S.,
Harris, B., Aikin, R., Doggett, J., Mayes, R., Keating, L.,
Wheeler, P. R., Parkhill, J., Barrell, B. G., Cole, S. T., Gordon, S. V. and
Hewinson, G.
The complete genome sequence of Mycobacterium bovis
Online Publication
PNAS 10.1073/pnas.1130426100 (Microbiology)
REFERENCE 2 (bases 1 to 343050)
AUTHORS
Direct Submission
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium bovis sequencing teams, TS Research Group, Veterinary
Laboratories Agency Weybridge, Woodham Lane, New Haw, Addlestone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA, UK. PT4 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France
Location/Qualifiers
1. 343050
/organism="Mycobacterium bovis subsp. bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
1. 1524
/gene="dnaA"
/locus_tag="Mb00001"
1. 1524
/gene="dnaA"
/locus_tag="Mb00001"
/note="Mb00001, dnaA, len: 507 aa. Equivalent to Rv0001,
len: 507 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.6% identity in 507 aa overlap). dnaA, chromosomal
replication initiator protein (see citations below),
equivalent to other Mycobacterial CHROMOSOMAL REPLICATION
INITIATOR PROTEINS e.g. P46388|DNAA MYCLE from
Mycobacterium leprae (502 aa); Q917L7|DNAA MYCPA from
Mycobacterium paratuberculosis (509 aa); P49990|DNAA_MYCAV

```

from Mycobacterium avium (508 aa); P49992|DNAA_MYCSM from
Mycobacterium smegmatis (504 aa); etc. Also highly similar
to others except in N-terminus e.g. Q9ZH75|DNAA_STRCH
CHROMOSOMAL REPLICATION INITIATOR PROTEIN from
Streptomyces chrysomallus (624 aa); Q9ZH76|DNAA_STRPE from
Streptomyces reticuli (643 aa); DNAA_ECOLI|P03004|B3702
chromosomal replication initiator protein from Escherichia
coli strain K12 (467 aa), FASTA scores: opt: 986, E(): 0,
(43.2% identity in 389 aa overlap); etc. Contains PS00017
ATP/GTP-binding site motif A (P-loop) and PS01008 Dnaa
protein signature. BELONGS TO THE DNAA FAMILY. Note that
the first base of this gene has been taken as base 1 of
the Mycobacterium bovis genomic sequence."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CHROMOSOMAL REPLICATION INITIATOR PROTEIN DNAA"
/protein_id="CAD92863.1"
/db_xref="GI:31616763"
/translation="MTDDPGSGTTVMNAVVSSELGDPKVDGSPSDANLSAPLTPQQ
RAWNLVQPLTIVEGFALLSVSPSSFVQNEIERHLRAPITDALSRRHQIQIGVRIAP
PATDEADDTVPSPENPATSPDTTDDNDIDDSAAAGDNGHSPSYFTEPRNTDS
ATAGVTSLSNRVTFTDFVIGASNRPAHAALAIAPARAYNPLFMGESGIGKTHLL
HAAGNYAQLRFFPMRVKYSYTFEFTNDPNSLRDDRKVPKRSYRDVDVLLVDQFPI
EKEGIGQEFFHTNTHANKQIVISSDRPKQALILSDRDRTPFEGLLIDVQFPE
LETRIALRKKQMERLAIPDDVLELIASSIERNIRELEGALIRVTAFASLNKTPIDK
ALAEIVLRDLIADANTMOISAAITMAATAYEDFTTVEELRGPKTRALAAQSRQIANYL
CRELTLSLPKIQAGFRDHTTVMVAQRKILSEMAERREVRDHFVKELTIRQRSKR"
2052..3260
/gene="dnaN"
/locus_tag="Mb00002"
2052..3260
/gene="dnaN"
/locus_tag="Mb00002"
/EC_number="2.7.7.7"
/notes="Mb00002, dnaN, len: 402 aa. Equivalent to Rv0002,
len: 402 aa from Mycobacterium tuberculosis strain H37Rv,
(99.8% identity in 402 aa overlap). dnaN, DNA polymerase
III (beta chain) (EC 2.7.7.7) (see citations below),
equivalent to other Mycobacterial DNA POLYMERASES III BETA
CHAIN e.g. NP_301130.1|NC_002677 from Mycobacterium leprae
(399 aa); Q9L7L6|DP3B_MYCPA from Mycobacterium avium
subsp. paratuberculosis (399 aa); P52851|DP3B_MYCSM from
Mycobacterium smegmatis (397 aa); etc. Also highly similar
to others e.g. P27903|DP3B_STRCO DNA POLYMERASE III BETA
CHAIN from Streptomyces coelicolor (376 aa), FASTA scores:
opt: 1189, E(): 0, (52.8% identity in 337 aa overlap);
P21174|DP3B_MICU from Micrococcus luteus (310 aa);
P52023|DP3B_SYN7 from Synecococcus sp. strain PCC 7942
(375 aa); etc. Overlaps and extends CDS in neighbouring
cosmid MTCY10H4.01."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="DNA POLYMERASE III (BETA CHAIN) DNAN (DNA
NUCLEOTIDYLTRANSFERASE)"
/protein_id="CAD92864.1"
/db_xref="GI:31616764"
/translation="MDAATIRVGLTDLTRLLRESFADAVSWAKNLPARPAPVPLSG
VILTGDNLGTSIGDFYVSAEQVGAETVSPGSLVSGRLSDITRALPNKPVGVHY
EGNRVALTCGNARFSLTPEVDYPTPLTPEETGLLPAELFAEAIQVIAAGRDDT
LPMLTGIRVEILGETVLAATDRTFRLAVRLKWSASSPDIEAAVLVPAKTLAEAAKAG
IGGSDVRLSGLTGPVGKGLLIGSGKRSSTRLLDAPFPKROLLPTEHTVATMD
VAELIEAIKLVADVADRGAVQVMEFADGSRVLSAGADDVGRAEELVDVYAGEPLTIA
FNPTYLDGLSLRSERVFEGTTACKPALLRPVSGDDRPVAGLNGGFPFPAVSTDYV
YLLMPVRLPFG"
3280..4437
/gene="recF"
/locus_tag="Mb00003"
3280..4437
/gene="recF"
/locus_tag="Mb00003"
/notes="Mb00003, recF, len: 385 aa. Equivalent to Rv0003,

```

```

len: 385 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.5% identity in 385 aa overlap). recF, DNA replication
and repair protein (see citations below), equivalent to
others Mycobacterial DNA replication and repair proteins
e.g. NP_301131.1|NC_002677 from Mycobacterium leprae (385
aa); Q9L7L5|RECF_MYCPA from Mycobacterium avium subsp.
paratuberculosis (385 aa); P50916|RECF_MYCSM from
Mycobacterium smegmatis (384 aa); etc. Also highly similar
to others e.g. P36176|RECF_STRCO DNA REPLICATION AND
REPAIR PROTEIN from Streptomyces coelicolor (373 aa);
NP_440892.1|NC_000911 from Synecococcus sp. strain PCC
6803 (384 aa); NP_469352.1|NC_003212 from Listeria innocua
(370 aa); etc. Contains PS00017 ATP/GTP-binding site motif
A (P-loop), PS00617 RecF protein signature 1, and PS00618
RecF protein signature 2. BELONGS TO THE RECF FAMILY."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="DNA REPLICATION AND REPAIR PROTEIN RECF
(SINGLE-STRAND DNA BINDING PROTEIN)"
/protein_id="CAD92865.1"
/db_xref="GI:31616765"
/translation="MYVRHLGLRDFRSWACVDLELHPGRTVFVGPNGYKTNLLEALW
YSTTLGSHRVSADEPLIRVCTDRAVISTIVNDGRECAVDLEIATGRVKNALNRSSV
RSTDVGVSVLAVLFAPEDLGLVKGDPADRRYRDDLAIVRPAIAAVRAEYERVRQ
RTALLKSPVGARYRGDGVFTLEWDSRLAEGAEVLAARIDVNLQAEVVKAYQL
LAPRSASIGYRASMDVTPGSEQSDTDLQALAAELAAARDEALERGVCUVGPH
RDDILIRLGQOPAKGFASGHEAWSLAVLAALYQLLRVDGEGPEVLDDVFAELDM
RRRALATAASAEQVLVTAVLEIDIPAGWDARRVHIDVADDDTGSMSVLP"
4434..4997
/locus_tag="Mb00004"
4434..4997
/locus_tag="Mb00004"
/notes="Mb00004, -, len: 187 aa. Equivalent to Rv0004, len:
187 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.5% identity in 187 aa overlap). Conserved hypothetical
protein (see citation below), highly similar, but longer
21 aa in N-terminus, to AAF33696.1|AF222789 unknown
protein from Mycobacterium avium subsp. paratuberculosis
(166 aa); and highly similar to NP_301132.1|NC_002677
conserved hypothetical protein from Mycobacterium leprae
(189 aa); S70990 hypothetical protein from Mycobacterium
smegmatis (194 aa). Also similar to in C-terminus to
C-terminal part of P35925|YREG_STRCO HYPOTHETICAL 19.8 KDA
PROTEIN (IN RECF-GYRB INTERGENIC REGION) from Streptomyces
coelicolor (190 aa), FASTA scores: opt: 404, E(): 3.9e-16,
(40.7% identity in 189 aa overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD92866.1"
/db_xref="GI:31616766"
/translation="MTGSDVRPDQNRGRLMKSPGLDVRRLDEAARARQDAG
RGRVASVAGRVAGRRRSWSGPDIRDQPLGKAARELAKKRGWSRVRAEGVILGW
SAVGHQIAEHARPTALNDGVLISVAESTAWATQLRIMQAQLAKIAAAGVNDVRSIL
KITGPAAPSVKRPRIHAGRPRTYG"
5123..7267
/gene="gyrB"
/locus_tag="Mb00005"
5123..7267
/gene="gyrB"
/locus_tag="Mb00005"
/EC_number="5.99.1.3"
/notes="Mb00005, gyrB, len: 714 aa. Equivalent to Rv0005,
len: 714 aa, from Mycobacterium tuberculosis strain H37Rv,
(99.9% identity in 714 aa overlap). gyrB, DNA gyrase
subunit B (EC 5.99.1.3) (see citations below), equivalent,
except in N-terminus, to other Mycobacterial DNA GYRASES
SUBUNIT B e.g. T10005 from Mycobacterium leprae (697 aa);
Q9L7L3|GYRB_MYCPA from Mycobacterium avium subsp.
paratuberculosis (677 aa) (has its N-terminus shorter);
P48555|GYRB_MYCSM from Mycobacterium smegmatis (675 aa);
etc. Also highly similar to others e.g. T10969 from

```

Streptomyces coelicolor (686 aa); P50075|GYBS_STRSH from

Query Match 100.0%; Score 26; DB 1; Length 343050;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GATGACCGATCGTCGGGGCGGGAAC 26

DB 241621 GATGACCGATCGTCGGGGCGGGAAC 241646

RESULT 14

BD273782

LOCUS

BD273782 498 bp DNA linear PAT 17-JUL-2003

DEFINITION Sequences nucleic acid from polypeptides exportes from mycobacteries, vector the comprenant and applications at diagnostic and the prevention from the tuberculose.

ACCSSION

BD273782.1 GI:33083550

VERSION

JP 2002534956-A/6.

KEYWORDS

Mycobacterium tuberculosis

SOURCE

Mycobacterium tuberculosis

ORGANISM

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium

tuberculosis complex.

1 (bases 1 to 498)

Gicquel, B., Portnoi, D., Lim, E., Pelicic, V., Guigueno, A. and

Salmoniere, Y. G. D. L.

Sequences nucleic acid from polypeptides exportes from

mycobacteries, vector the comprenant and applications at diagnostic

and the prevention from the tuberculose

Patent: JP 2002534956-A 6 22-OCT-2002;

INSTITUT PASTEUR

OS Mycobacterium tuberculosis

PN JP 2002534956-A/6

PD 22-OCT-2002

PF 14-AUG-1998 JP 2000509849

PR 14-AUG-1997 FR 97/10404, 11-SEP-1997 FR 97/11325 PI

BRIGITTE GICQUEL, DENIS PORTNOI, ENG-MONG LIM, VLADIMIR PELICIC, PI

AGNES GUIGUENO, YVES GOGUET DE LA SALMONIERE

PC C12N15/09, A61K39/04, A61K48/00, A61P11/00, A61P31/04, C07K14/35,

PC C07K16/12,

PC C07K19/00, C12N1/21, C12P21/02, C12Q1/02, C12Q1/68, G01N33/50, PC

G01N33/53//

PC C12N15/09, C12R1:32), C12N15/00, (C12N15/00, C12R1:32) CC seq

ID no. 1F

FH Key

FT CDS

Location/Qualifiers

(4). (495).

1. .498

/organism="Mycobacterium tuberculosis"

/mol_type="genomic DNA"

/db_xref="taxon:1773"

Query Match 92.3%; Score 24; DB 6; Length 498;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 TGACGGATCGTCGGGGCGGGAAC 26

DB 1 TGACGGATCGTCGGGGCGGGAAC 24

RESULT 15

AX004914

LOCUS

AX004914 498 bp DNA linear PAT 24-AUG-2000

DEFINITION

Sequence 33 from Patent WO9909186.

ACCESSION

AX004914

VERSION

AX004914.1 GI:9928297

KEYWORDS

Mycobacterium tuberculosis

SOURCE

Mycobacterium tuberculosis

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CDS

ORIGIN

Query Match

Best Local Similarity

Matches

QY

DB

Search completed:

Job time :

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:01:30 ; Search time 2050.13 Seconds
(without alignments)
378.717 Million cell updates/sec

Title: US-10-624-714-3

Perfect score: 26

Sequence: 1 gatgacgatctcgggggcggaac 26

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:**

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_hci:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: em_gss_hum:**

18: em_gss_inv:**

19: em_gss_pin:**

20: em_gss_vrt:**

21: em_gss_fun:**

22: em_gss_mam:**

23: em_gss_mus:**

24: em_gss_pro:**

25: em_gss_rod:**

26: em_gss_phg:**

27: em_gss_vrl:**

28: gb_gsl1:**

29: gb_gsl2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20.4	78.5	468	14	CF486281
C 2	20.4	78.5	624	13	CA072458
C 3	20.4	78.5	656	13	CA077580
C 4	20.4	78.5	667	14	CA294341

C 5	20.4	78.5	730	13	CA120257
C 6	20.4	78.5	809	14	CA206115
C 7	19.8	76.2	321	14	CF880400
C 8	19.4	74.6	990	29	CNSQ3V80
C 9	19.4	74.6	1015	29	CNSQ4CW1
C 10	19.2	73.8	697	14	CF441948
C 11	19.2	73.8	738	29	AG041131
C 12	19.2	73.8	855	12	BI772667
C 13	19.2	73.8	969	28	BH996777
C 14	19.2	73.8	1130	13	BQ682710
C 15	18.6	71.5	380	12	BJ184864
C 16	18.6	71.5	417	28	BH612923
C 17	18.6	71.5	446	10	BF888444
C 18	18.6	71.5	691	12	BM593760
C 19	18.6	71.5	692	12	BM48618
C 20	18.6	71.5	725	12	BM36163
C 21	18.6	71.5	802	28	BZ530710
C 22	18.6	71.5	897	29	CG061380
C 23	18.6	71.5	943	29	CNSQ1XQX
C 24	18.6	71.5	1009	29	CG061377
C 25	18.6	71.5	1269	12	BM912093
C 26	18.4	70.8	391	13	BY004072
C 27	18.2	70.0	386	12	BI190155
C 28	18.2	70.0	494	14	CA708796
C 29	18.2	70.0	495	13	BU981235
C 30	18.2	70.0	507	12	BU473736
C 31	18.2	70.0	530	13	BU976388
C 32	18.2	70.0	573	12	BJ465343
C 33	18.2	70.0	629	14	CB351131
C 34	18.2	70.0	651	12	BJ468586
C 35	18.2	70.0	879	29	CG235838
C 36	18.2	70.0	914	10	BF134217
C 37	18.2	70.0	971	12	BI949119
C 38	18	69.2	168	28	BH416883
C 39	18	69.2	171	28	BH230615
C 40	18	69.2	196	13	BQ100937
C 41	18	69.2	212	28	BH229021
C 42	18	69.2	214	28	BH229373
C 43	18	69.2	246	28	BH229023
C 44	18	69.2	248	28	BH229343
C 45	18	69.2	253	28	BH228902

ALIGNMENTS

RESULT 1
CF486281/c
LOCUS
DEFINITION
POLI_36_G06_g1_A002 Pollen Sorghum bicolor cDNA clone
POLI_36_G06_A002 5', mRNA sequence.
ACCESSION
CF486281
VERSION
CF486281.1 GI:34515150
KEYWORDS
EST.
SOURCE
Sorghum bicolor (sorghum)
ORGANISM
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
1 (bases 1 to 468)
Cordonnier-Pratt, M.-M., Suzuki, Y., Sugano S., Klein, R.R., Liang, C., Sun, F., Sullivan, R., Eastman, A. and Pratt, L.H.
AUTHORS
An EST database from Sorghum: pollen
TITLE
Unpublished (2003)
JOURNAL
Other ESTs: POLI_36_G05.bl_A002
COMMENT
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in

the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.

Seq primer: Sug5 (CTTCTGCTCTAAAGATGCG).

FEATURES

source

```
1..468
Location/Qualifiers
  /organism="Sorghum bicolor"
  /mol_type="mRNA"
  /cultivar="BTx623"
  /db_xref="taxon:4559"
  /clone="POL1 36 G06 A002"
  /lab_host="DH10B-T1 phage-resistant E. coli"
  /clone_lib="Pollen"
  /note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; the library was prepared from polyA+ RNA from pollen at the late vacuolated-vacuolated stage of development. Pollen was harvested from greenhouse-grown panicles of sorghum line BTx623. Panicles were removed from the flag leaf prior to emergence, when no detectable anylase is present in pollen of male-fertile lines. This stage represents pollen collected from anthers about 8-14 days prior to anthesis. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5'-prime DraIII site is CACTGTGTG, 3'-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."
```

ORIGIN

```
Query Match      78.5%; Score 20.4; DB 14; Length 468;
Best Local Similarity 95.5%; Pred. No. 2.9e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GATGACGGATCGTCGGGGCGG 22
|||||
Db 299 GATGACGGATCGTCGGGGCGG 278
|||||
```

RESULT 2

```
CA072458/c
LOCUS SCEPAM1016D06.g AM1 Saccharum officinarum cDNA clone SCEPAM1016D06
DEFINITION 5', mRNA sequence.
ACCESSION CA072458
VERSION CA072458.1 GI:34924609
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
```

```
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Saccharum.
```

```
1 (bases 1 to 624)
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
```

The libraries that made SUCEST

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica

Universidade Estadual de Campinas

Caixa Postal 6010, 13083-970, Campinas SP, Brazil

Tel: 55 19 3788 1137

Fax: 55 19 3788 1089

Email: parruda@unicamp.br

Clone distribution: clone distribution information can be found

through the Brazilian Clone Collection Center (BCCC) at

http://www.bcccenter.fcav.unesp.br

Plate: 016 row: D column: 06

Seq primer: T7 Promoter Primer.

FEATURES

source

```
1..624
Location/Qualifiers
  /organism="Saccharum officinarum"
```

```
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCEPAM1016D06"
/lab_host="DH10B"
/clone_lib="AM1"
```

/note="Organ: Apical meristem and tissues surrounding of mature plants; Vector: pSport1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from [Apical meristem and tissues surrounding of mature plants]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <http://sucest.lad.ic.unicamp.br/public>"

ORIGIN

```
Query Match      78.5%; Score 20.4; DB 13; Length 624;
Best Local Similarity 95.5%; Pred. No. 2.9e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 GATGACGGATCGTCGGGGCGG 22
|||||
Db 384 GATGACGGATCGTCGGGGCGG 363
|||||
```

RESULT 3

CA077580/c

```
LOCUS SCQSAM1033B02.g AM1 Saccharum officinarum cDNA clone SCQSAM1033B02
DEFINITION 5', mRNA sequence.
ACCESSION CA077580
VERSION CA077580.1 GI:34929852
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
```

```
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Saccharum.
```

1 (bases 1 to 656)

Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.

The libraries that made SUCEST

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica

Universidade Estadual de Campinas

Caixa Postal 6010, 13083-970, Campinas SP, Brazil

Tel: 55 19 3788 1137

Fax: 55 19 3788 1089

Email: parruda@unicamp.br

Clone distribution: clone distribution information can be found

through the Brazilian Clone Collection Center (BCCC) at

http://www.bcccenter.fcav.unesp.br

Plate: 033 row: B column: 02

Seq primer: T7 Promoter Primer.

FEATURES

source

```
1..656
Location/Qualifiers
  /organism="Saccharum officinarum"
  /mol_type="mRNA"
  /db_xref="taxon:4547"
/clone="SCQSAM1033B02"
/lab_host="DH10B"
/clone_lib="AM1"
```

/note="Organ: Apical meristem and tissues surrounding of mature plants; Vector: pSport1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from [Apical meristem and tissues surrounding of mature plants]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and

1.5 Kb were directionally cloned into the vector. Details

of each source of RNA and library construction can be obtained at <http://sucet.lad.ic.unicamp.br/public>

```

ORIGIN
Query Match      78.5%; Score 20.4; DB 13; Length 656;
Best Local Similarity 95.5%; Pred. No. 2.9e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGG 22
    |||||
Db 402 GATGACGGGTCGTCGGGGCGG 381
    |||||

RESULT 4
CA294341/c
LOCUS
DEFINITION
SCSGLV1011E02.g LV1 Saccharum officinarum cDNA clone SCSGLV1011E02
5', mRNA sequence.
ACCESSION
CA294341
VERSION
CA294341.1 GI:36056600
KEYWORDS
EST.
SOURCE
Saccharum officinarum
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.
1 (bases 1 to 667)
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
The libraries that made SUCEST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
Plate: 011 row: E column: 02
Seq primer: T7 Promoter Primer.
Location/Qualifiers
1..667
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCSGLV1011E02"
/lab_host="DH10B"
/clone_lib="LV1"
/note="Organ: Etiolated leaves from in vitro grown
seedlings; Vector: pSport1; Site_1: SalI; Site_2: NotI; An
unidirectional cDNA library generated from [etiolated
leaves from in vitro grown seedlings]. cDNA was prepared
from polyA+ mRNA using SuperScript Plasmid System Kit
(Invitrogen). The double-strand cDNAs were fractionated
in a sepharose CL-2B 40cm-columns and fragments sizing
between 0.8 and 1.5 Kb were directionally cloned into the
vector. Details of each source of RNA and library
construction can be obtained at
http://sucet.lad.ic.unicamp.br/public

ORIGIN
Query Match      78.5%; Score 20.4; DB 14; Length 667;
Best Local Similarity 95.5%; Pred. No. 2.9e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGG 22
    |||||
Db 437 GATGACGGGTCGTCGGGGCGG 416
    |||||

RESULT 5

```

```

CA120257/c
LOCUS
DEFINITION
SCCCLR1075D10.g LR1 Saccharum officinarum cDNA clone SCCCLR1075D10
5', mRNA sequence.
ACCESSION
CA120257
VERSION
CA120257.1 GI:34973565
KEYWORDS
EST.
SOURCE
Saccharum officinarum
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.
1 (bases 1 to 730)
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
The libraries that made SUCEST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
Plate: 075 row: D column: 10
Seq primer: T7 Promoter Primer.
Location/Qualifiers
1..730
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCCCLR1075D10"
/lab_host="DH10B"
/clone_lib="LR1"
/note="Organ: Leaf roll from field grown adult plants
(large insert library); Vector: pSport1; Site_1: SalI;
Site_2: NotI; An unidirectional cDNA library generated
from [leaf roll from field grown adult plants (large
insert library)]. cDNA was prepared from polyA+ mRNA
using SuperScript Plasmid System Kit (Invitrogen). The
double-strand cDNAs were fractionated in a sepharose
CL-2B 40cm-columns and fragments sizing between 0.8 and
1.5 Kb were directionally cloned into the vector. Details
of each source of RNA and library construction can be
obtained at http://sucet.lad.ic.unicamp.br/public

ORIGIN
Query Match      78.5%; Score 20.4; DB 13; Length 730;
Best Local Similarity 95.5%; Pred. No. 2.9e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGG 22
    |||||
Db 421 GATGACGGGTCGTCGGGGCGG 400
    |||||

RESULT 6
CA206115/c
LOCUS
DEFINITION
SCVPFL1073E12.g FL1 Saccharum officinarum cDNA clone SCVPFL1073E12
5', mRNA sequence.
ACCESSION
CA206115
VERSION
CA206115.1 GI:35245100
KEYWORDS
EST.
SOURCE
Saccharum officinarum
ORGANISM
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.
1 (bases 1 to 809)
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.

```

TITLE JOURNAL COMMENT

The libraries that made SUCEST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parada@unicamp.br
Clone distribution: Clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
Plate: 073 row: E column: 12
Seq primer: T7 Promoter Primer.
Location/Qualifiers
1. .809

FEATURES source

/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCVPL1073E12"
/lab_host="DH10B"
/clone_lib="FL1"
/note="Organ: Inflorescence at beginning of development
(1cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI;
An unidirectional cDNA library generated from
inflorescence at beginning of development (1cm-long).
cDNA was prepared from polyA+ mRNA using SuperScript
Plasmid System Kit (Invitrogen). The double-strand cDNAs
were fractionated in a sepharose CL-2B 40cm-columns and
fragments sizing between 0.8 and 1.5 Kb were
directionally cloned into the vector. Details of each
source of RNA and library construction can be obtained at
http://sucest.lad.ic.unicamp.br/public"

ORIGIN

Query Match 78.5%; Score 20.4; DB 14; Length 809;
Best Local Similarity 95.5%; Pred. No. 2.9e+03;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGGGG 22
|||||
Db 387 GATGACGGGTCGTCGGGGGGG 366

RESULT 7 CF880400/c

LOCUS trico22xf24.g1 T.reesei mycelial culture, Version 6 October 2003
DEFINITION Hypocrea jecorina cDNA clone trico22xf24, mRNA sequence.

ACCESSION VERSION

CF880400.1 GI:38135082

KEYWORDS SOURCE

ORGANISM Hypocrea jecorina (anamorph: Trichoderma reesei)
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
1 (bases 1 to 321)
Diener, S.E., Dankmeyer, L., Dunn-Coleman, N., Houfek, T.D.,
Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M. and
Dean, R.A.

REFERENCE AUTHORS

Analysis of the protein processing and secretion pathways in a
Trichoderma reesei EST dataset
Unpublished (2003)
Contact: Ralph A. Dean

JOURNAL COMMENT

Fungal Genomics Laboratory
North Carolina State University
Campus Box 7251, Raleigh, NC 27695, USA
Tel: 919-513-0020
Fax: 919-513-0024
Email: ralph.dean@ncsu.edu
Seq primer: Tr-R1 primer.
Location/Qualifiers
1. .321

FEATURES source

/organism="Hypocrea jecorina"

/mol_type="mRNA"

/db_xref="taxon:4547"

/clone="SCVPL1073E12"

/lab_host="DH10B"

/clone_lib="FL1"

/note="Organ: Inflorescence at beginning of development
(1cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI;
An unidirectional cDNA library generated from
inflorescence at beginning of development (1cm-long).
cDNA was prepared from polyA+ mRNA using SuperScript
Plasmid System Kit (Invitrogen). The double-strand cDNAs
were fractionated in a sepharose CL-2B 40cm-columns and
fragments sizing between 0.8 and 1.5 Kb were
directionally cloned into the vector. Details of each
source of RNA and library construction can be obtained at
http://sucest.lad.ic.unicamp.br/public"

/organism="Hypocrea jecorina"

/mol_type="mRNA"

/db_xref="taxon:51453"

/clone="trico22xf24"

/dev_stage="mycelia"

/clone_lib="T.reesei mycelial culture, Version 6 October 2003"

/note="Vector: pREP3; Site_1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Query Match 76.2%; Score 19.8; DB 14; Length 321;
Best Local Similarity 91.3%; Pred. No. 4.7e+03;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGGGG 23
|||||
Db 153 GATGACGGATCGTCGGGGGGG 131

RESULT 8

LOCUS CNS03V80

DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 061H01 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION VERSION

AL262089.1 GI:7983715

KEYWORDS GSS; genome survey sequence.

SOURCE ORGANISM

Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.
1
Roest Crolius H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
PUBMED 10835645

REFERENCE AUTHORS

Roest Crolius H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
Estimate of human gene number provided by genome-wide analysis
using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)

TITLE JOURNAL MEDLINE PUBMED

Roest Crolius H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
PUBMED 10899143

REFERENCE AUTHORS

Roest Crolius H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,
Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

TITLE JOURNAL MEDLINE PUBMED

Roest Crolius H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,
Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
PUBMED 10899143

REFERENCE AUTHORS

Roest Crolius H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,
Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
PUBMED 10899143

TITLE JOURNAL MEDLINE PUBMED

Roest Crolius H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,
Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
PUBMED 10899143

REFERENCE AUTHORS

Roest Crolius H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,
Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
PUBMED 10899143

TITLE JOURNAL MEDLINE PUBMED

Roest Crolius H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,
Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
PUBMED 10899143

REFERENCE AUTHORS

Roest Crolius H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,
Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
PUBMED 10899143

Roest Crolius H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,
Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
PUBMED 10899143

Roest Crolius H., Jaillon, O., Dasilva, C., Ozouf-Costaz, C.,
Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
PUBMED 10899143

PUC-Ori"

ORIGIN

Query Match 74.6%; Score 19.4; DB 29; Length 990;
 Best Local Similarity 95.2%; Pred. No. 6.2e+03;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GACGGATCGTCGGGGCGGGA 24
 |||||
 DB 54 GACGGACCGTCGGGGCGGGA 74
 |||||

RESULT 9

CNS04CW1 1015 bp DNA linear GSS 01-SEP-2000
 LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
 DEFINITION 101G22 of library G from Tetraodon nigroviridis, genomic survey
 sequence.

ACCESSION AL284986
 VERSION AL284986.1 GI:8023392
 KEYWORDS GSS; genome survey sequence.

SOURCE

ORGANISM Tetraodon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
 Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE

1 Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
 Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
 Saurin,W. and Weissenbach,J.

TITLE Estimate of human gene number provided by genome-wide analysis
 using Tetraodon nigroviridis DNA sequence

JOURNAL

MEDLINE Nat. Genet. 25 (2), 235-238 (2000)
 PUBMED 20296633
 PUBMED 10835645

REFERENCE

2 Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
 Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
 Saurin,W., Bernot,A. and Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the
 freshwater pufferfish Tetraodon nigroviridis

JOURNAL

MEDLINE Genome Res. 10 (7), 939-949 (2000)
 PUBMED 20359837
 PUBMED 10899143

REFERENCE

3 (bases 1 to 1015)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
 Bp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)

COMMENT This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the tetraodon nigroviridis
 genome. For more information, please take a look at
 http://www.genoscope.cns.fr/Tetraodon.

FEATURES

Location/Qualifiers
 source
 1..1015
 /organism="Tetraodon nigroviridis"
 /mol_type="genomic DNA"
 /db_xref="taxon:99883"
 /clone="101G22"
 /clone_lib="G"
 /note="Genoscope sequence ID : C0BG101BD11LP1-end : T7"

ORIGIN

Query Match 74.6%; Score 19.4; DB 29; Length 1015;
 Best Local Similarity 95.2%; Pred. No. 6.2e+03;
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GACGGATCGTCGGGGCGGGA 24
 |||||
 DB 85 GACGGACCGTCGGGGCGGGA 105
 |||||

RESULT 10

CF441948/c 697 bp mRNA linear EST 04-SEP-2003
 LOCUS EST678293 normalized cDNA library of onion Allium cepa cDNA clone
 DEFINITION ACAF045, mRNA sequence.

ACCESSION

VERSION CF441948
 KEYWORDS CF441948.1 GI:34464650
 SOURCE EST.

SOURCE

ORGANISM Allium cepa (onion)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
 Allium.

REFERENCE

1 (bases 1 to 697)
 Havey,M.J., Cheung,F., Van Aken,S., Utterback,T. and Town,C.D.
 TITLE Expressed Sequence Tags from a normalized library of mixed onion
 tissues (Allium cepa)

JOURNAL

COMMENT Unpublished (2003)

COMMENT

Contact: Havey MJ
 Department of Horticulture
 USDA-ARS and University of Wisconsin
 1575 Linden Drive, Madison, WI 53706, USA

Tel: 608-262-1830
 Fax: 608-262-4743

Email: mjhavey@facstaff.wisc.edu

TIGR sequence name ACAF045.STR. For more information:

http://haveylab.hort.wisc.edu

Seq primer: CAG GAA ACA GCT ATG ACC.

FEATURES

Location/Qualifiers
 source
 1..697
 /organism="Allium cepa"
 /mol_type="mRNA"
 /cultivar="Red Creole(bulbs), unknown(callus), Ebano &
 Texas legend(roots)"
 /db_xref="taxon:4679"
 /clone="ACAF045"
 /tissue_type="Callus, roots, and young bulbs"
 /clone_lib="normalized cDNA library of onion"
 /note="Vector: pCMVSPORT6.1-cdb (Invitrogen); Site 1:
 EcoRV (5'); Site 2: NotI (3'); Equal molar amounts of mRNA
 from callus, roots, and young bulbs were combined to
 synthesize the library. Normalization to enrich for
 low-copy transcripts was performed by proprietary
 techniques of Invitrogen."

ORIGIN

Query Match 73.8%; Score 19.2; DB 14; Length 697;
 Best Local Similarity 87.5%; Pred. No. 7.4e+03;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TGACGGATCGTCGGGGCGGGAAC 26
 |||||

DB 40 TGTCGGATCGTCGGGGCCAGGAC 17
 |||||

RESULT 11

AG041131/c 738 bp DNA linear GSS 01-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-018124.F, genomic survey sequence.

ACCESSION

VERSION AG041131
 KEYWORDS AG041131.1 GI:16569856
 SOURCE GSS.

SOURCE

ORGANISM Pan troglodytes (chimpanzee)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

REFERENCE

1 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
 Totoki,Y., Watanabe,H. and Sakaki,Y.

TITLE BAC end sequences of library PTB

JOURNAL

REFERENCE Unpublished
 2 (bases 1 to 738)

AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
 Toki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
 1-7-22 Suehiro-chou, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: chimps@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the R&D process and may have higher chance of
 clone tracking errors.

PRIMERS

Sequencing: -21M13

LIBRARY

Vector : PKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

1. .738

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

/clone="PTB-018L24.F"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN

Query Match 73.8%; Score 19.2; DB 29; Length 738;
 Best Local Similarity 87.5%; Pred. No. 7.4e+03;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGA 24

Db 446 GTTAGGATCGTCGGGGCGGGA 423

RESULT 12
 BI772667 855 bp mRNA linear EST 25-SEP-2001
 LOCUS 603058158F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5207391 5',
 DEFINITION mRNA sequence.

ACCESSION BI772667

VERSION BI772667.1 GI:15764245

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 855)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@bbs-r@mail.nih.gov

Tissue Procurement: Life Technologies, Inc.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM11520 row: d column: 16

High quality sequence start: 26

High quality sequence stop: 819.

Location/Qualifiers

1. .855

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5207391"

/lab_host="DH10B"

/clone_lib="NIH_MGC_122"

FEATURES

source

ORIGIN

Query Match 73.8%; Score 19.2; DB 28; Length 969;
 Best Local Similarity 87.5%; Pred. No. 7.3e+03;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGA 24

Db 612 GATGACGGATCGGAGGGGAGGGA 589

/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
 Site 1: NotI; Site 2: EcoRV (destroyed); RNA source
 anonymous pool of 24 week female lung, 16 week female
 spleen, and 20-22 week male spleens. Library is oligo-dT
 primed and directionally cloned (EcoRV site is destroyed
 upon cloning). Average insert size 1.4 kb, insert size
 range 1-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 026. Note:
 this is a NIH_MGC Library."

ORIGIN

Query Match 73.8%; Score 19.2; DB 12; Length 855;
 Best Local Similarity 87.5%; Pred. No. 7.3e+03;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TGACGGATCGTCGGGGCGGGAAC 26

Db 21 TTACGGATCGTCGGGGCGGGAAC 44

RESULT 13

BI899677/c

LOCUS

DEFINITION

OS00744 Ostreococcus tauri genomic shotgun library Ostreococcus

tauri genomic clone otb01g07.b 5', genomic survey sequence.

ACCESSION BI899677

VERSION BI899677.1 GI:22551172

KEYWORDS GSS.

SOURCE Ostreococcus tauri

ORGANISM Ostreococcus tauri

REFERENCE Eukaryota; Viridiplantae; Chlorophyta; Prasinophyceae; Mamiellales;

Mamiellaceae; Ostreococcus.

1 (bases 1 to 969)

Derelle, E., Ferraz, C., Legoda, P., Bychenie, S., Cooke, R., Regad, P.,

Sabau, X., Courties, C., Delseny, M., Demaille, J., Picard, A. and

Moreau, H.

(Chlorophytae, Prasinophyceae): the genome of Ostreococcus tauri

cell

J. Phycol. 38 (6), 1150-1156 (2002)

Contact: Moreau H

Laboratoire Arago

CNRS UMR 7628

BP 44, Avenue Fontaule, 66651 Banyuls sur mer, France

Tel: (33) 468887309

Fax: (33) 468887398

Email: h.moreau@obs-banyuls.fr

Seq primer: reverse

Class: Shotgun.

Location/Qualifiers

1. .969

/organism="Ostreococcus tauri"

/mol_type="genomic DNA"

/strain="OTTH0595"

/db_xref="taxon:70448"

/clone="otb01g07.b"

/clone_lib="Ostreococcus tauri genomic shotgun library"

/notes="Vector: Bluescript; Site 1: EcoRV; Site 2: EcoRV;

Shotgun library prepared after sonication of the genomic

DNA. Blunt ligation in EcoRV site of Bluescript. Size

selection of the inserts after agarose electrophoresis

between 1 and 3 kb."

ORIGIN

Query Match 73.8%; Score 19.2; DB 28; Length 969;
 Best Local Similarity 87.5%; Pred. No. 7.3e+03;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGA 24

Db 612 GATGACGGATCGGAGGGGAGGGA 589

Protonemata were blended by the POLYTRON, and then cultivated on the BCDATG medium for 13- 14 days under the continuous light.

```

Location/Qualifiers
1. 380
/organism="Physcomitrella patens subsp. patens"
/mol_type="mRNA"
/sub_species="patens"
/db_xref="taxon:145481"
/clone="pphb5e07"
/tissue_type="mixture of chloronemata, caulonemata and
malformed buds"
/clone_lib="normalized full length cDNA library,
chloronemata, caulonemata and malformed buds"

```

leted: April 29, 2004, 11:36:50
2054.13 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 06:08:26 ; Search time 485.333 Seconds

(without alignments)
242.328 Million cell updates/sec

Title: US-10-624-714-3

Perfect score: 26

Sequence: 1 gatgacgagtcgctggggcggaac 26

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTU5_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.6	75.4	469	13	US-10-412-699B-1270 Sequence 1270, Ap
C 2	19.6	75.4	469	16	US-10-374-780A-983 Sequence 983, App
C 3	19.6	75.4	996	13	US-10-617-038-85 Sequence 85, Appl
C 4	19.2	73.8	722	13	US-10-282-122A-36752 Sequence 36752, A
C 5	18.6	71.5	1376	15	US-10-237-551-217 Sequence 217, App
C 6	18.6	71.5	1599	15	US-10-121-988-144 Sequence 144, App
C 7	18.6	71.5	1599	15	US-10-200-562-144 Sequence 144, App
C 8	18.6	71.5	1599	15	US-10-237-551-144 Sequence 144, App
C 9	18.6	71.5	1599	15	US-10-237-551-218 Sequence 218, App
C 10	18.6	71.5	1599	15	US-10-237-551-245 Sequence 245, App
C 11	18.6	71.5	1623	15	US-10-121-988-85 Sequence 85, Appl
C 12	18.6	71.5	1623	15	US-10-200-562-85 Sequence 85, Appl
C 13	18.6	71.5	1623	15	US-10-237-551-85 Sequence 85, Appl
C 14	18.6	71.5	4817	16	US-10-311-388-2 Sequence 2, Appl

C 15	18.6	71.5	8933	16	US-10-311-388-1 Sequence 1, Appl
C 16	18.6	71.5	154746	10	US-09-827-688-8 Sequence 8, Appl
C 17	18	69.2	413	13	US-10-027-632-141517 Sequence 141517, Appl
C 18	18	69.2	413	16	US-10-027-632-141517 Sequence 141517, Appl
C 19	18	69.2	1510	16	US-10-264-049-928 Sequence 928, App
C 20	18	69.2	42999	9	US-09-799-462A-17 Sequence 17, Appl
C 21	18	69.2	42999	10	US-09-836-911A-17 Sequence 17, Appl
C 22	18	69.2	42999	14	US-09-738-630-73 Sequence 73, Appl
C 23	18	69.2	42999	14	US-10-125-767-17 Sequence 17, Appl
C 24	18	69.2	42999	15	US-10-151-081-17 Sequence 17, Appl
C 25	18	69.2	42999	15	US-10-287-313-17 Sequence 17, Appl
C 26	18	69.2	42999	15	US-10-219-694-17 Sequence 17, Appl
C 27	17.8	68.5	1392	16	US-10-369-493-41455 Sequence 41455, A
C 28	17.8	68.5	2186	13	US-10-425-114-27260 Sequence 27260, A
C 29	17.8	68.5	2199	13	US-10-430-011-122 Sequence 122, App
C 30	17.6	67.7	386	13	US-10-276-774-181 Sequence 181, App
C 31	17.6	67.7	574	13	US-10-282-122A-24043 Sequence 24043, A
C 32	17.6	67.7	1007	13	US-10-282-122A-19220 Sequence 19220, A
C 33	17.6	67.7	1014	9	US-09-815-242-6207 Sequence 6207, Ap
C 34	17.6	67.7	1014	9	US-09-815-242-9682 Sequence 9682, Ap
C 35	17.6	67.7	1014	13	US-10-282-122A-20511 Sequence 20511, A
C 36	17.6	67.7	1014	13	US-10-282-122A-39301 Sequence 39301, A
C 37	17.6	67.7	2392	10	US-09-813-432-23 Sequence 23, Appl
C 38	17.6	67.7	2392	13	US-10-246-583-23 Sequence 23, Appl
C 39	17.6	67.7	2392	16	US-10-174-364-23 Sequence 23, Appl
C 40	17.6	67.7	3772	9	US-09-989-722-61 Sequence 61, Appl
C 41	17.6	67.7	3772	9	US-09-989-723-61 Sequence 61, Appl
C 42	17.6	67.7	3772	9	US-09-989-729-61 Sequence 61, Appl
C 43	17.6	67.7	3772	9	US-09-989-727-61 Sequence 61, Appl
C 44	17.6	67.7	3772	9	US-09-989-731-61 Sequence 61, Appl
C 45	17.6	67.7	3772	9	US-09-989-732-61 Sequence 61, Appl

ALIGNMENTS

RESULT 1

US-10-412-699B-1270/C
; Sequence 1270, Application US/10412699B
; Publication No. US20040045049A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: Zhang, James
; APPLICANT: Fromm, Michael E.
; APPLICANT: Heard, Jacqueline E.
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc J.
; APPLICANT: Brown, Pierre E.
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James S.
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Samaha, Raymond R.
; APPLICANT: Pilgrim, Marsha L.
; APPLICANT: Creselman, Robert A.
; APPLICANT: DuBell Arnold N.
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Kumamoto, Roderick
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MB1-0048CIP
; CURRENT APPLICATION NUMBER: US/10/412,699B
; CURRENT FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: 09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 09/489,376
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: 09/506,720
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 09/533,030
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,392

; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,029
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/532,591
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,648
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/713,994
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: 09/819,142
; PRIOR FILING DATE: 2001-03-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2011
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1270
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (405)..(405)
; OTHER INFORMATION: n is a, c, g, or t
US-10-412-699B-1270

Query Match 75.4%; Score 19.6; DB 13; Length 469;
Best Local Similarity 84.6%; Pred. No. 21;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGGAC 26
Db 431 GACGACGGCTCGTCGGGGCGGGC 406

RESULT 2

US-10-374-780A-983/c
; Sequence 983, Application US/10374780A
; Publication No. US2004001992A1

; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Heard, Jacqueline E
; APPLICANT: Haake, Volker
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Keddle, James
; APPLICANT: Broun, Pierre E
; APPLICANT: Pilgrim, Marsha L
; APPLICANT: Dubell III, Arnold T
; APPLICANT: Pineda, Omaira
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 CIP
; CURRENT APPLICATION NUMBER: US/10/374,780A
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 09/837,944
; PRIOR FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 60/310,847
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: 09/934,455
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/336,049
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/338,692
; PRIOR FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: 10/171,468
; PRIOR FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 10/225,066
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: 10/225,067
; PRIOR FILING DATE: 2002-08-09

; PRIOR APPLICATION NUMBER: 10/225,068
; PRIOR FILING DATE: 2002-08-09
; NUMBER OF SEQ ID NOS: 2906
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 983
; LENGTH: 469
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (405)..(405)
; OTHER INFORMATION: n is a, c, g, or t
US-10-374-780A-983

Query Match 75.4%; Score 19.6; DB 16; Length 469;
Best Local Similarity 84.6%; Pred. No. 21;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGGAC 26
Db 431 GACGACGGCTCGTCGGGGCGGGC 406

RESULT 3

US-10-617-038-85/c
; Sequence 85, Application US/10617038
; Publication No. US20040057963A1

; GENERAL INFORMATION:
; APPLICANT: Andersen, Peter
; APPLICANT: Rosenkrands, Ida
; APPLICANT: Stryhn, Anette
; TITLE OF INVENTION: Therapeutic TB Vaccine
; FILE REFERENCE: SSI5AUSA
; CURRENT APPLICATION NUMBER: US/10/617,038
; PRIOR FILING DATE: 2003-07-11
; PRIOR APPLICATION NUMBER: DK PA 2002 01098
; PRIOR FILING DATE: 2002-07-13
; PRIOR APPLICATION NUMBER: US 60/401,725
; PRIOR FILING DATE: 2002-08-07
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 85
; LENGTH: 996
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-617-038-85

Query Match 75.4%; Score 19.6; DB 13; Length 996;
Best Local Similarity 84.6%; Pred. No. 19;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGGAC 26
Db 602 GATGGGGCTCGTTGGGGCGGGGAC 577

RESULT 4

US-10-282-122A-36752
; Sequence 36752, Application US/10282122A
; Publication No. US20040029129A1

; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 36752
; LENGTH: 722
; TYPE: DNA
; ORGANISM: Salmonella paratyphi A
US-10-282-122A-36752

Query Match 73.8%; Score 19.2; DB 13; Length 722;
Best Local Similarity 87.5%; Pred. No. 30;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GATGACGGATCGTCGGGGCGGGA 24
Db 293 GATGACGGATCGTCGGGGCGGCTGGA 316

RESULT 5
US-10-237-551-217/c
; Sequence 217, Application US/10237551
; Publication No. US20030165820A1
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121-538C3
; CURRENT APPLICATION NUMBER: US/10/237,551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 217
; LENGTH: 1376
; TYPE: DNA
; ORGANISM: Herpes simplex virus
US-10-237-551-217

Query Match 71.5%; Score 18.6; DB 15; Length 1376;
Best Local Similarity 84.0%; Pred. No. 51;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 ATGACGGATCGTCGGGGCGGGAAC 26
Db 443 ACGTCGGTTCGTCGGGGCGGGAAC 419

RESULT 6

US-10-121-988-144/c
; Sequence 144, Application US/10121988
; Publication No. US20030068327A1
; GENERAL INFORMATION:
; APPLICANT: Hosken, Nancy Ann
; APPLICANT: McGowan, Patrick
; APPLICANT: Sleath, Paul R.
; APPLICANT: Mossman, Sally P.
; APPLICANT: Evans, Lawrence S.
; APPLICANT: Swanson, Ryan M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121-538C1
; CURRENT APPLICATION NUMBER: US/10/121,988
; CURRENT FILING DATE: 2002-04-11
; NUMBER OF SEQ ID NOS: 183
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144
; LENGTH: 1599
; TYPE: DNA
; ORGANISM: HSV2
US-10-121-988-144

Query Match 71.5%; Score 18.6; DB 15; Length 1599;
Best Local Similarity 84.0%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 ATGACGGATCGTCGGGGCGGGAAC 26
Db 1162 ACGTCGGTTCGTCGGGGCGGGAAC 1138

RESULT 7

US-10-200-562-144/c
; Sequence 144, Application US/10200562
; Publication No. US20030165819A1
; GENERAL INFORMATION:
; APPLICANT: McGowen, Patrick
; APPLICANT: Hosken, Nancy A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121-538C2
; CURRENT APPLICATION NUMBER: US/10/200,562
; CURRENT FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 144
; LENGTH: 1599
; TYPE: DNA
; ORGANISM: HSV2
US-10-200-562-144

Query Match 71.5%; Score 18.6; DB 15; Length 1599;
Best Local Similarity 84.0%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 ATGACGGATCGTCGGGGCGGGAAC 26
Db 1162 ACGTCGGTTCGTCGGGGCGGGAAC 1138

RESULT 8

US-10-237-551-144/c
; Sequence 144, Application US/10237551
; Publication No. US20030165820A1
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND

```

/ / TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
/ / FILE REFERENCES: Z10121.58C3
/ / CURRENT APPLICATION NUMBER: US/10/237,551
/ / CURRENT FILING DATE: 2002-09-06
/ / NUMBER OF SEQ ID NOS: 254
/ / SOFTWARE: FastSEQ for Windows Version 4.0
/ / SEQ ID NO 144
/ / LENGTH: 1599
/ / TYPE: DNA
/ / ORGANISM: HSV2
/ / US-10-237-551-144

```

```
Query Match      71.5%; Score 18.6; DB 15; Length 1599;
Best Local Similarity 84.0%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

Qy

2 ATGACGGATCGTCGGGGCGGGAAC 26
| | | | | | | | | |
Dβ

1162 ACGTCGGTTCGTCTGGGCGCGGGAAC 1138

```

RESULT 9
US-10-237-551-218/c
/ Sequence 218, Application US/10237551
/ Publication NO. US20030165820A1
/ GENERAL INFORMATION:
/ APPLICANT: Day, Craig H.
/ APPLICANT: Hosken, Nancy A.
/ APPLICANT: Parsons, Joseph M.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
/ TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
/ FILE REFERENCE: 210121.536C3
/ CURRENT APPLICATION NUMBER: US/10/237,551
/ CURRENT FILING DATE: 2002-03-06
/ NUMBER OF SEQ ID NOS: 254
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 218
/ LENGTH: 1599
/ TYPE: DNA
/ ORGANISM: Herpes simplex virus
US-10-237-551-218

```

```
Query Match      71.5%; Score 18.6; DB 15; Length 1599;
Best Local Similarity 84.0%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

Qy 2 ATGACGGATCGTCGGGGCGGGAAC 26
db 1162 ACGTCGGTTCGTGGGCGCGGGAAC 1138

```

RESULT 10
US-10-237-551-245/c
; Sequence 245, Application US/10237551
; Publication No. US20030165820A1
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C3
; CURRENT APPLICATION NUMBER: US/10/237,551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 245
; LENGTH: 1599
; TYPE: DNA
; ORGANISM: Herpes simplex virus
US-10-237-551-245

```

Query Match 71.5%; Score 18.6; DB 15; Length 1599;

```

Best Local Similarity 84.0%; Pred. NO. 50;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      2 ATGACGGATCGTCGGGGCGGGAAC 26
          | | | | | | | | | |
Db      1162 AGTCGGTTCGTCGGGCGGGGAAC 1138

```

```

RESULT 11
US-10-121-988-85/c
/ Sequence 85, Application US/10121988
/ Publication No. US20030068327A1
/ GENERAL INFORMATION:
/ APPLICANT: Hosken, Nancy Ann
/ APPLICANT: McGowan, Patrick
/ APPLICANT: Sleath, Paul R.
/ APPLICANT: Mossman, Sally P.
/ APPLICANT: Evans, Lawrence S.
/ APPLICANT: Swanson, Ryan M.
/ APPLICANT: McNeill, Patricia D.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
/ TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
/ FILE REFERENCE: 210121.538C1
/ CURRENT APPLICATION NUMBER: US/10/121.988
/ CURRENT FILING DATE: 2002-04-11
/ NUMBER OF SEQ ID NOS: 183
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 85
/ LENGTH: 1623
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-121-988-85

```

Query Match	71.5%;	Score 18.6;	DB 15;	Length 1623;
Best Local Similarity	84.0%;	Pred. No. 50;		
Matches	21;	Conservative 0;	Mismatches 4;	Indels 0;
Gaps	0;			
QY	2	ATGACGGATCGTCGGGGCGGGAAC	26	
Db	1183	ACGTGGGTTCGTGGGGCGGGAAC	1159	

```

RESULT 12
US-10-200-562-85/c
; Sequence 85, Application US/10200562
; Publication No. US20030165819A1
; GENERAL INFORMATION:
; APPLICANT: McGowen, Patrick
; APPLICANT: Hosken, Nancy A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C2
; CURRENT APPLICATION NUMBER: US/10/200,562
; CURRENT FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 85
; LENGTH: 1623
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-200-562-85

```

Query Match	71.5%	Score 18.6	DB 15	Length 1623
Best Local Similarity	84.0%	Pred. No. 50		
Matches 21	Conservative 0	Mismatches 4	Indels 0	Gaps 0
QY	2	ATGACGGATCGTCGGGGCGGGAAC	26	
Db	1183	AGGTCGGTTCGTCGGGGCGGGAAC	1159	

RESULT 13
US-10-237-551-85/c

; Sequence 85, Application US/10237551
; Publication No. US20030165820A1
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210421-538C3
; CURRENT APPLICATION NUMBER: US/10/237,551
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 85
; LENGTH: 1623
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-237-551-85

Query Match 71.5%; Score 18.6; DB 15; Length 1623;
Best Local Similarity 84.0%; Pred. No. 50;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 ATGACGGATCGTCGGGGCGGGAC 26
Db 1183 ACCTCGGTCGTCGGGGCGGGAC 1159

RESULT 14

US-10-311-388-2/c
; Sequence 2, Application US/10311388
; Publication No. US20030215831A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Sanicola-Nadel, Michele
; APPLICANT: Hession, Catherine
; APPLICANT: Tizard Jr., Richard
; APPLICANT: Bonventure, Joseph
; TITLE OF INVENTION: RENAL REGULATORY ELEMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 00689-502-061 (BGN-2)
; CURRENT APPLICATION NUMBER: US/10/311,388
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: USSN 60/212,131
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4817
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-311-388-2

Query Match 71.5%; Score 18.6; DB 16; Length 4817;
Best Local Similarity 84.0%; Pred. No. 43;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGAA 25
Db 1026 GATGACGGATGTCGGGGCGGGAA 1002

RESULT 15

US-10-311-388-1/c
; Sequence 1, Application US/10311388
; Publication No. US20030215831A1
; GENERAL INFORMATION:
; APPLICANT: Biogen, Inc.
; APPLICANT: Sanicola-Nadel, Michele
; APPLICANT: Hession, Catherine
; APPLICANT: Tizard Jr., Richard
; APPLICANT: Bonventure, Joseph
; TITLE OF INVENTION: RENAL REGULATORY ELEMENTS AND METHODS OF USE THEREOF
; FILE REFERENCE: 00689-502-061 (BGN-2)

; CURRENT APPLICATION NUMBER: US/10/311,388
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: USSN 60/212,131
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 8933
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-311-388-1

Query Match 71.5%; Score 18.6; DB 16; Length 8933;
Best Local Similarity 84.0%; Pred. No. 40;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GATGACGGATCGTCGGGGCGGGAA 25
Db 4821 GATGACGGATGTCGGGGCGGGAA 4797

Search completed: April 29, 2004, 20:43:48
Job time : 485.333 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:25:04 ; Search time 452.706 Seconds
(without alignments)
2872.264 Million cell updates/sec

Title: US-10-624-714-4

Perfect score: 30

Sequence: 1 gcaagcggagggatcagctactgcaaca 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rnd:*
36: em_htg_man:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	30	100.0	411	6	BD273781	BD273781 Sequences
C 2	30	100.0	411	6	AX004912	AX004912 Sequences
C 3	30	100.0	498	6	BD273782	BD273782 Sequences
C 4	30	100.0	498	6	AX004914	AX004914 Sequences
C 5	30	100.0	959	6	BD273780	BD273780 Sequences
C 6	30	100.0	959	6	AX004906	AX004906 Sequences
C 7	30	100.0	960	6	BD273779	BD273779 Sequences
C 8	30	100.0	960	6	AX004895	AX004895 Sequences
C 9	30	100.0	961	6	BD273778	BD273778 Sequences
C 10	30	100.0	961	6	AX004889	AX004889 Sequences
C 11	30	100.0	1063	6	BD273783	BD273783 Sequences
C 12	30	100.0	1063	6	AX004916	AX004916 Sequences
C 13	30	100.0	1243	6	BD273777	BD273777 Sequences
C 14	30	100.0	1243	6	AX004882	AX004882 Sequences
C 15	30	100.0	14664	1	AE006931	AE006931 Mycobacte
C 16	30	100.0	341957	15	EX842572	EX842572 Mycobacte
C 17	30	100.0	343050	1	EX248334	EX248334 Mycobacte
C 18	21.2	70.7	1008	1	AF411068	AF411068 Spingomo
C 19	21	70.0	259550	1	EX294136	EX294136 Pirellula
C 20	20.4	68.0	152264	2	AC119572	AC119572 Ustilago
C 21	20.4	68.0	152831	2	AC021743	AC021743 Homo sapi
C 22	20.4	68.0	156569	2	AC110190	AC110190 Homo sapi
C 23	20.4	68.0	164314	9	AC091691	AC091691 Homo sapi
C 24	20.2	67.3	208050	1	AL646083	AL646083 Ralstonia
C 25	20	66.7	1881	3	AF164027	AF164027 Leishmani
C 26	20	66.7	29001	2	AC019796	AC019796 Drosophil
C 27	20	66.7	175167	3	AC093101	AC093101 Drosophil
C 28	20	66.7	265413	3	AE003646	AE003646 Drosophil
C 29	20	66.7	316756	3	DROSADH06	AE003412 Drosophil
C 30	19.6	65.3	11576	1	AE001994	AE001994 Deinococc
C 31	19.6	65.3	83885	2	AC122729	AC122729 Medicago
C 32	19.6	65.3	133077	8	AP005103	AP005103 Oryza sat
C 33	19.6	65.3	153592	9	AL163533	AL163533 Human DNA
C 34	19.6	65.3	161906	2	AC027206	AC027206 Homo sapi
C 35	19.6	65.3	300800	1	AP005036	AP005036 Streptomy
C 36	19.6	65.3	327861	2	AC092922	AC092922 Homo sapi
C 37	19.4	64.7	792	1	AF462398	AF462398 Serratia
C 38	19.4	64.7	792	1	AF462399	AF462399 Enterobac
C 39	19.4	64.7	876	1	AB038771	AB038771 Escherich
C 40	19.4	64.7	876	1	AF325133	AF325133 Klebsiell
C 41	19.4	64.7	876	1	AF325134	AF325134 Klebsiell
C 42	19.4	64.7	876	1	AY143430	AY143430 Klebsiell
C 43	19.4	64.7	876	1	AY156923	AY156923 Escherich
C 44	19.4	64.7	941	1	AF311345	AF311345 Escherich
C 45	19.4	64.7	1220	1	AF252623	AF252623 Klebsiell

ALIGNMENTS

RESULT 1
BD273781/c
LOCUS BD273781
DEFINITION Sequences nucleic acid from polypeptides exportes from mycobacteries, vector the complement and applications at diagnostic and the prevention from the tuberculose.
ACCESSION BD273781
VERSION BD273781.1 GI:33083549
KEYWORDS JP 2002534956-A/5.
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 411)

Pred. No. is the number of results predicted by chance to have a

AUTHORS Gicquel,B., Portnoi,D., Lim,E., Pelicic,V., Guigueno,A. and Salmoniere,Y.G.D.L.

TITLE Sequences nucleic acid from polypeptides exportes from mycobacteries, vector the comprenant and applications at diagnostic and the prevention from the tuberculose

JOURNAL Patent: JP 2002534956-A 5 22-OCT-2002; INSTITUT PASTEUR

COMMENT OS Mycobacterium tuberculosis
PN JP 2002534956-A/5
PD 22-OCT-2002
PF 14-AUG-1998 JP 2000509849
PR 14-AUG-1997 FR 97/10404,11-SEP-1997 FR 97/11325 PI
BRIGITTE GICQUEL,DENIS PORTNOI,ENG-MONG LIM,VLADIMIR PELICIC, PI
AGNES GUIGUENO,YVES GOGUET DE LA SALMONIERE
PC C12N15/09,A61K39/04,A61K48/00,A61P11/00,A61P31/04,C07K14/35,
PC C07K16/12,
PC C07K19/00,C12N1/21,C12P21/02,C12Q1/02,C12Q1/68,G01N33/50, PC
G01N33/53//
PC (C12N15/09,C12R1:32),C12N15/00,(C12N15/00,C12R1:32) CC seq
ID no. 1D
FH Key Location/Qualifiers
FT CDS (1)..(408).

FEATURES
source 1..411
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"

ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 411;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAACGGAGGGCGATCAGTACTGCCAACA 30
Db 64 GCAACGGAGGGCGATCAGTACTGCCAACA 35

RESULT 2
AX004912/c
LOCUS AX004912 411 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 31 from Patent WO9909186.
ACCESSION AX004912
VERSION AX004912.1 GI:9928295
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE
1 Portnoi,D. and Guigueno,A.
AUTHORS
TITLE Polypeptide nucleic sequences exported from mycobacteria, vectors comprising same and uses for diagnosing and preventing tuberculosis
JOURNAL Patent: WO 9909186-A 31 25-FEB-1999;
PORTNOI DENIS (FR); GUIGUENO AGNES (FR)

FEATURES
source 1..411
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"
1..411
/notes="unnamed protein product; seq ID no 1D"
/codon_start=1
/transl_table=11
/protein_id="CAC04930.1"
/db_xref="GI:9928295"
/db_xref="RENTREMBL:CAC04930"
/translation="MKTGATTTRELLAVLIALPFGAVALLAEPFATGASDPCAAS
EVARTVGSVAKSGDYLDGHPETNOMVTAVLQQVQPGSVASLKAHPFANPKVASDLH
ALSQPLTDLSTRSLPISGLQIGLMQAVQGAR"

ORIGIN

Query Match 100.0%; Score 30; DB 6; Length 411;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAACGGAGGGCGATCAGTACTGCCAACA 30
Db 64 GCAACGGAGGGCGATCAGTACTGCCAACA 35

RESULT 3
BD273782/c
LOCUS BD273782 498 bp DNA linear PAT 17-JUL-2003
DEFINITION Sequences nucleic acid from polypeptides exportes from mycobacteries, vector the comprenant and applications at diagnostic and the prevention from the tuberculose.

ACCESSION BD273782
VERSION BD273782.1 GI:33083550
KEYWORDS JP 2002534956-A/6.
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE
1 (bases 1 to 498)
AUTHORS Gicquel,B., Portnoi,D., Lim,E., Pelicic,V., Guigueno,A. and Salmoniere,Y.G.D.L.
TITLE Sequences nucleic acid from polypeptides exportes from mycobacteries, vector the comprenant and applications at diagnostic and the prevention from the tuberculose
JOURNAL Patent: JP 2002534956-A 6 22-OCT-2002;
INSTITUT PASTEUR

COMMENT OS Mycobacterium tuberculosis
PN JP 2002534956-A/6
PD 22-OCT-2002
PF 14-AUG-1998 JP 2000509849
PR 14-AUG-1997 FR 97/10404,11-SEP-1997 FR 97/11325 PI
BRIGITTE GICQUEL,DENIS PORTNOI,ENG-MONG LIM,VLADIMIR PELICIC, PI
AGNES GUIGUENO,YVES GOGUET DE LA SALMONIERE
PC C12N15/09,A61K39/04,A61K48/00,A61P11/00,A61P31/04,C07K14/35,
PC C07K16/12,
PC C07K19/00,C12N1/21,C12P21/02,C12Q1/02,C12Q1/68,G01N33/50, PC
G01N33/53//
PC (C12N15/09,C12R1:32),C12N15/00,(C12N15/00,C12R1:32) CC seq
ID no. 1F
FH Key Location/Qualifiers
FT CDS (4)..(495).

FEATURES
source 1..498
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"

ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 498;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAACGGAGGGCGATCAGTACTGCCAACA 30
Db 151 GCAACGGAGGGCGATCAGTACTGCCAACA 122

RESULT 4
AX004914/c
LOCUS AX004914 498 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 33 from Patent WO9909186.
ACCESSION AX004914
VERSION AX004914.1 GI:9928297
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium

```

tuberculosis complex.
1
REFERENCE
AUTHORS Portnoi,D. and Guigueno,A.
TITLE Polypeptide nucleic sequences exported from mycobacteria, vectors
JOURNAL comprising same and uses for diagnosing and preventing tuberculosis
PATENT: WO 9909186-A 33 25-FEB-1999;
PORTNOI DENIS (FR); GUIGUENO AGNES (FR)
FEATURES
source Location/Qualifiers
1..498
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"
<4..498
/note="unnamed protein product; seq ID no 1F"
/codon_start=1
/transl_table=11
/protein_id="CAC04931.1"
/db_xref="GI:9928298"
/db_xref="RENTREMBL:CAC04931"
translation="RTVGGNSQAGPSOTRRKPVANRKESSMKTGTATRRRLAVL
IALALPGAVALLAEPASGASDPCASAEVATVGSVAKSMGDIYLDSPETNVMYTA
LQQVGPVGSVSLKHFANPKVADLHLSQPLDLSRCSLPISGLQIGLMQAVQ
GARR"
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 498;
Best Local Similarity 100.0%; Pred. No. 0.16; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0;
QY 1 GCAACGCGAGGGCGATCAGTACTGCCAACA 30
|||||
DB 151 GCAACGCGAGGGCGATCAGTACTGCCAACA 122
|||||

RESULT 5
BD273780/c
LOCUS 959 bp DNA linear PAT 17-JUL-2003
DEFINITION Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the comprenant and applications at diagnostic
and the prevention from the tuberculose.
ACCESSION BD273780.1 GI:33083548
VERSION JP 2002534956-A/4.
KEYWORDS Mycobacterium tuberculosis
SOURCE Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 498;
Best Local Similarity 100.0%; Pred. No. 0.16; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0;
QY 1 GCAACGCGAGGGCGATCAGTACTGCCAACA 30
|||||
DB 151 GCAACGCGAGGGCGATCAGTACTGCCAACA 122
|||||

RESULT 6
BD273779/c
LOCUS 959 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 25 from Patent WO9909186.
ACCESSION AX004906
VERSION AX004906.1 GI:9928293
KEYWORDS Mycobacterium tuberculosis
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 959;
Best Local Similarity 100.0%; Pred. No. 0.14; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0;
QY 1 GCAACGCGAGGGCGATCAGTACTGCCAACA 30
|||||
DB 722 GCAACGCGAGGGCGATCAGTACTGCCAACA 693
|||||

RESULT 7
BD273779/c
LOCUS 960 bp DNA linear PAT 17-JUL-2003
DEFINITION Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the comprenant and applications at diagnostic
and the prevention from the tuberculose.
ACCESSION BD273779.1 GI:33083547
VERSION JP 2002534956-A/3.
KEYWORDS Mycobacterium tuberculosis
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 959;
Best Local Similarity 100.0%; Pred. No. 0.14; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0;
QY 1 GCAACGCGAGGGCGATCAGTACTGCCAACA 30
|||||
DB 722 GCAACGCGAGGGCGATCAGTACTGCCAACA 693
|||||

tuberculosis complex.
1
REFERENCE
AUTHORS Portnoi,D. and Guigueno,A.
TITLE Polypeptide nucleic sequences exported from mycobacteria, vectors
JOURNAL comprising same and uses for diagnosing and preventing tuberculosis
PATENT: WO 9909186-A 33 25-FEB-1999;
PORTNOI DENIS (FR); GUIGUENO AGNES (FR)
FEATURES
source Location/Qualifiers
1..498
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"
<4..498
/note="unnamed protein product; seq ID no 1F"
/codon_start=1
/transl_table=11
/protein_id="CAC04931.1"
/db_xref="GI:9928298"
/db_xref="RENTREMBL:CAC04931"
translation="RTVGGNSQAGPSOTRRKPVANRKESSMKTGTATRRRLAVL
IALALPGAVALLAEPASGASDPCASAEVATVGSVAKSMGDIYLDSPETNVMYTA
LQQVGPVGSVSLKHFANPKVADLHLSQPLDLSRCSLPISGLQIGLMQAVQ
GARR"
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 498;
Best Local Similarity 100.0%; Pred. No. 0.16; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0;
QY 1 GCAACGCGAGGGCGATCAGTACTGCCAACA 30
|||||
DB 151 GCAACGCGAGGGCGATCAGTACTGCCAACA 122
|||||

RESULT 5
BD273780/c
LOCUS 959 bp DNA linear PAT 17-JUL-2003
DEFINITION Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the comprenant and applications at diagnostic
and the prevention from the tuberculose.
ACCESSION BD273780.1 GI:33083548
VERSION JP 2002534956-A/4.
KEYWORDS Mycobacterium tuberculosis
SOURCE Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 498;
Best Local Similarity 100.0%; Pred. No. 0.16; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0;
QY 1 GCAACGCGAGGGCGATCAGTACTGCCAACA 30
|||||
DB 151 GCAACGCGAGGGCGATCAGTACTGCCAACA 122
|||||

RESULT 6
BD273779/c
LOCUS 959 bp DNA linear PAT 24-AUG-2000
DEFINITION Sequence 25 from Patent WO9909186.
ACCESSION AX004906
VERSION AX004906.1 GI:9928293
KEYWORDS Mycobacterium tuberculosis
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 959;
Best Local Similarity 100.0%; Pred. No. 0.14; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0;
QY 1 GCAACGCGAGGGCGATCAGTACTGCCAACA 30
|||||
DB 722 GCAACGCGAGGGCGATCAGTACTGCCAACA 693
|||||

RESULT 7
BD273779/c
LOCUS 960 bp DNA linear PAT 17-JUL-2003
DEFINITION Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the comprenant and applications at diagnostic
and the prevention from the tuberculose.
ACCESSION BD273779.1 GI:33083547
VERSION JP 2002534956-A/3.
KEYWORDS Mycobacterium tuberculosis
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 959;
Best Local Similarity 100.0%; Pred. No. 0.14; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0;
QY 1 GCAACGCGAGGGCGATCAGTACTGCCAACA 30
|||||
DB 722 GCAACGCGAGGGCGATCAGTACTGCCAACA 693
|||||
```

Salmoniere, Y.G.D.L.
Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the comprénant and applications at diagnostic
and the prevention from the tuberculose
Patent: JP 2002534956-A 3 22-OCT-2002;
INSTITUT PASTEUR
OS Mycobacterium tuberculosis
PN JP 2002534956-A/3
PD 22-OCT-2002
PF 14-AUG-1998 JP 2000509849 97/10404,11-SEP-1997 FR 97/11325 PI
BRIGITTE GICQUEL, DENIS PORTNOI, ENG-MONG LIM, VLADIMIR FELICIC, PI
AGNES GUIGUENO, YVES GOGUET DE LA SALMONIERE
PC C12N15/09, A61K39/04, A61K48/00, A61P11/00, A61P31/04, C07K14/35,
C07K16/12,
PC C07K19/00, C12N1/21, C12P21/02, C12Q1/02, C12Q1/68, G01N33/50, PC
G01N33/53//
PC (C12N15/09, C12R1:32), C12N15/00, (C12N15/00, C12R1:32) CC seq
ID no. 1A'
FH Key Location/Qualifiers
FT CDS (1)..(960).

FEATURES
source
1..960
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"

ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 960;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAACGGCGAGCGGATCAGTACTGCCAACA 30
Db 723 GCAACGGCGAGCGGATCAGTACTGCCAACA 694

RESULT 8
AX004895/c
LOCUS
DEFINITION
Sequence 14 from Patent WO9909186.
ACCESSION
AX004895
VERSION
AX004895.1 GI:9928291
KEYWORDS
Mycobacterium tuberculosis
SOURCE
Mycobacterium tuberculosis
ORGANISM
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE
1 Portnoi, D. and Guigueno, A.
Polypeptide nucleic sequences exported from mycobacteria, vectors
comprising same and uses for diagnosing and preventing tuberculosis
Patent: WO 9909186-A 14 25-FEB-1999;
PORTNOI DENIS (FR); GUIGUENO AGNES (FR)

FEATURES
source
1..960
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"

CDS
1..174
/notes="unnamed protein product; seq ID no 1B"
/codon_start=1
/transl_table=11
/protein_id="CAC04928.1"
/db_xref="GI:9928292"
/db_xref="RENTREMBL:CAC04928"
/translations="IADFAYSVAQLFGDVTGARVAGVQPRHHIRVRLRGLPGVGC
LRHPRQFAGYLSO"

ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 960;
Best Local Similarity 100.0%; Pred. No. 0.14;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAACGGCGAGCGGATCAGTACTGCCAACA 30
Db 723 GCAACGGCGAGCGGATCAGTACTGCCAACA 694

RESULT 9
BD273778/c
LOCUS
DEFINITION
Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the comprénant and applications at diagnostic
and the prevention from the tuberculose.
BD273778
ACCESSION
BD273778.1 GI:33083546
VERSION
BD273778
KEYWORDS
JP 2002534956-A/2.
SOURCE
Mycobacterium tuberculosis
ORGANISM
Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1 (Bases 1 to 961)
Gicquel, B., Portnoi, D., Lim, E., Felicic, V., Guigueno, A. and
Salmoniere, Y.G.D.L.
Sequences nucleic acid from polypeptides exportes from
mycobacteries, vector the comprénant and applications at diagnostic
and the prevention from the tuberculose
Patent: JP 2002534956-A 2 22-OCT-2002;
INSTITUT PASTEUR
OS Mycobacterium tuberculosis
PN JP 2002534956-A/2
PD 22-OCT-2002
PF 14-AUG-1998 JP 2000509849 97/10404,11-SEP-1997 FR 97/11325 PI
BRIGITTE GICQUEL, DENIS PORTNOI, ENG-MONG LIM, VLADIMIR FELICIC, PI
AGNES GUIGUENO, YVES GOGUET DE LA SALMONIERE
PC C12N15/09, A61K39/04, A61K48/00, A61P11/00, A61P31/04, C07K14/35,
C07K16/12,
PC C07K19/00, C12N1/21, C12P21/02, C12Q1/02, C12Q1/68, G01N33/50, PC
G01N33/53//
PC (C12N15/09, C12R1:32), C12N15/00, (C12N15/00, C12R1:32) CC seq
ID no. 1A'
FH Key Location/Qualifiers
FT CDS (1)..(960).

FEATURES
source
1..961
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"

ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 961;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAACGGCGAGCGGATCAGTACTGCCAACA 30
Db 724 GCAACGGCGAGCGGATCAGTACTGCCAACA 695

RESULT 10
AX004889/c
LOCUS
DEFINITION
Sequence 8 from Patent WO9909186.
ACCESSION
AX004889
VERSION
AX004889.1 GI:9928289
KEYWORDS
Mycobacterium tuberculosis
SOURCE
Mycobacterium tuberculosis
ORGANISM
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE
1

AUTHORS Portnoi, D. and Guigueno, A.
TITLE Polypeptide nucleic sequences exported from mycobacteria, vectors comprising same and uses for diagnosing and preventing tuberculosis
JOURNAL PORTNOI DENIS (FR); GUIGUENO AGNES (FR)
FEATURES
source 1. .961
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"
<13. .>197
/notes="unnamed protein product; seq id no 1A"
/codon_start=1
/transl_table=11
/protein_id="CAC04927.1"
/db_xref="GI:9928290"
/db_xref="REMTREMBL:CAC04927"
/translation="RLFGAAFWRRHMCRCRPAATPHTAGPVAGVGNWGVFPSP QAVRWLFVAVARRHCRCL"
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 961;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAACGCGAGGCGGATCAGTACTGCCAACCA 30
|||||
Db 724 GCAACGCGAGGCGGATCAGTACTGCCAACCA 695
|||||
RESULT 11
BD273783/c
LOCUS
DEFINITION Sequences nucleic acid from polypeptides exportees from mycobacteries, vector the complement and applications at diagnostic and the prevention from the tuberculose.
ACCESSION BD273783
VERSION BD273783.1 GI:33083551
KEYWORDS JP 2002534956-A/7.
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1 (bases 1 to 1063)
AUTHORS Gicquel, B., Portnoi, D., Lim, E., Pellicic, V., Guigueno, A. and Salmoniere, Y. G. D. L.
TITLE Sequences nucleic acid from polypeptides exportees from mycobacteries, vector the complement and applications at diagnostic and the prevention from the tuberculose
JOURNAL PATENT: JP 2002534956-A 7 22-OCT-2002;
INSTITUT PASTEUR
COMMENT OS Mycobacterium tuberculosis
PN JP 2002534956-A/7
PD 22-OCT-2002
PF 14-AUG-1998 JP 2000509849
FR 14-AUG-1997 FR 97/10404, 11-SEP-1997 FR 97/11325 PI
BRIGHT GICQUEL, DENIS PORTNOI, ENG-MONG LIM, VLADIMIR PELICIC, PI AGNES GUIGUENO, YVES COGUET DE LA SALMONIERE
PC C12N15/09, A61K39/04, A61K48/00, A61P11/00, A61P31/04, C07K14/35, PC C07K16/12,
PC C07K19/00, C12N1/21, C12N21/02, C12Q1/02, C12Q1/68, G01N33/50, PC G01N33/53//
PC (C12N15/09, C12R1:32), C12N15/00, (C12N15/00, C12R1:32) CC seq ID no. 2
FH key Location/Qualifiers
FT CDS (53) . . (82)
FT CDS (180) . . (437)
FT CDS (451) . . (858)
FT CDS (862) . . (1062).
FEATURES
source 1. .1063
/organism="Mycobacterium tuberculosis"

/mol_type="genomic DNA"
/db_xref="taxon:1773"
ORIGIN
Query Match 100.0%; Score 30; DB 6; Length 1063;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAACGCGAGGCGGATCAGTACTGCCAACCA 30
|||||
Db 514 GCAACGCGAGGCGGATCAGTACTGCCAACCA 485
|||||
RESULT 12
AX004916/c
LOCUS
DEFINITION Sequences 35 from Patent WO9909186.
ACCESSION AX004916
VERSION AX004916.1 GI:9928299
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE 1
AUTHORS Portnoi, D. and Guigueno, A.
TITLE Polypeptide nucleic sequences exported from mycobacteria, vectors comprising same and uses for diagnosing and preventing tuberculosis
JOURNAL PATENT: WO 9909186-A 35 25-FEB-1999;
PORTNOI DENIS (FR); GUIGUENO AGNES (FR)
FEATURES
source 1. .1063
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"
53. .85
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC04932.1"
/db_xref="GI:9928300"
/db_xref="REMTREMBL:CAC04932"
/translation="MRVELVVDQ"
180. .440
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC04933.1"
/db_xref="GI:9928301"
/db_xref="REMTREMBL:CAC04933"
/translation="MKTNQRHLRLRLRNINNCWISDPSYSRAVRAREPTEDRVHAFG VDRTPAGVGGAEGRDRMTDRRGRLPGRTRVNPVNSQTRKP"
451. .861
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC04934.1"
/db_xref="GI:9928302"
/db_xref="REMTREMBL:CAC04934"
/translation="MKTGTTATRRRLVLTALPGAVALLAEPATGASDPCAAS EVATVGSVAKSMGDYLDSPETNQMTAVLQCVGPGSVASLKAHFENPKVADLH ALSQPLDLSTRCSLPSGLQIGLMOVQARR"
862. .51062
/notes="unnamed protein product; seq ID no2"
/codon_start=1
/transl_table=11
/protein_id="CAC04935.1"
/db_xref="GI:9928303"
/db_xref="REMTREMBL:CAC04935"
/translation="MPDRRRVRRSRREAAVAYRGVPPSGRRSGVAGPGVGVSTGS SQGVPCGMVTSRRFSGVGSRSRL"
ORIGIN

```

Query Match      100.0%; Score 30; DB 6; Length 1063;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAACGGAGGGGATCAGTACTGCGCAACA 30
Db 514 GCAACGGAGGGGATCAGTACTGCGCAACA 485

RESULT 13
BD273777/c
LOCUS
DEFINITION
Sequences nucleic acid from polypeptides exported from
mycobacteries, vector the complement and applications at diagnostic
and the prevention from the tuberculose.
BD273777
BD273777.1 GI:33083545
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE
AUTHORS
Gicquel,B., Portnoi,D., Lim,B., Pelicic,V., Guigueno,A. and
Salmoniere,Y.G.D.L.
TITLE
Sequences nucleic acid from polypeptides exportees from
mycobacteries, vector the complement and applications at diagnostic
and the prevention from the tuberculose
JOURNAL
Patent: JP 2002534956-A 1 22-OCT-2002;
INSTITUT PASTEUR
COMMENT
OS Mycobacterium tuberculosis
PN JP 2002534956-A/1
PD 22-OCT-2002
PF 14-AUG-1998 JP 2000509849
PR 14-AUG-1997 FR 97/10404, 11-SEP-1997 FR 97/11325 PI
BRIGITTE GICQUEL, DENIS PORTNOI, ENG-MONG LIM, VLADIMIR PELICIC, PI
AGNES GUIGUENO, YVES GOGUST DE LA SALMONIERE
PC C12N15/09, A61K39/04, A61K48/00, A61P11/00, A61P31/04, C07K14/35,
PC C07K16/12,
PC C07K19/00, C12N1/21, C12P21/02, C12Q1/02, C12Q1/68, G01N33/50, PC
G01N33/53//
PC (C12N15/09, C12R1:32), C12N15/00, (C12N15/00, C12R1:32) CC seq
ID no. 1
FH key Location/Qualifiers
FT CDS (21)..(38)
FT CDS (124)..(453)
FT CDS (473)..(481)
FT CDS (573)..(671)
FT CDS (719)..(928)
FT CDS (942)..(1241).

FEATURES
source
1. .1243
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"

ORIGIN
Query Match      100.0%; Score 30; DB 6; Length 1243;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAACGGAGGGGATCAGTACTGCGCAACA 30
Db 1005 GCAACGGAGGGGATCAGTACTGCGCAACA 976

RESULT 14
AX004882/c
LOCUS
DEFINITION
Sequence 1 from Patent WO9909186.
ACCESSION
AX004882

```

```

VERSION
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis
Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE
AUTHORS
Portnoi,D. and Guigueno,A.
TITLE
Polypeptide nucleic sequences exported from mycobacteria, vectors
comprising same and uses for diagnosing and preventing tuberculosis
Patent: WO 9909186-A 1 25-FEB-1999;
PORTNOI DENIS (FR); GUIGUENO AGNES (FR)
JOURNAL
Location/Qualifiers
FEATURES
source
1. .1243
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"
21..41
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC04921.1"
/db_xref="GI:9928283"
/db_xref="RENTREMBL:CAC04921"
/translation="MVVGM"
124..456
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC04922.1"
/db_xref="GI:9928284"
/db_xref="RENTREMBL:CAC04922"
/translation="MAGQSELRFDVPLTLAEASRYLVVPRATLATWADGVERSPA
NAPAVQCOPIAFDAISVAQLFQDVTGARVAGVQQRHHIRFVRLGVLGGVGLRHP
QFAGYLSQ"
473..484
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC04923.1"
/db_xref="GI:9928285"
/db_xref="RENTREMBL:CAC04923"
/translation="MSW"
573..674
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC04924.1"
/db_xref="GI:9928286"
/db_xref="RENTREMBL:CAC04924"
/translation="MTATRLRNRLDSTASSPGKPPALTATNP"
719..931
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC04925.1"
/db_xref="GI:9928287"
/db_xref="RENTREMBL:CAC04925"
/translation="MNCWISDPSYRAVRAEPTEDRVHAFGVDRTPAGVGGAEGRDG
RMDRRRELPGRRTVANPQTRKP"
942..1241
/note="unnamed protein product; seq ID no1"
/codon_start=1
/transl_table=11
/protein_id="CAC04926.1"
/db_xref="GI:9928288"
/db_xref="RENTREMBL:CAC04926"
/translation="MKTGTATRRRLAVIALALPGAAVALLAEPSATGSDPCAAS
EVATGVSVAKSGDYLDSPHETNQVMTAVLQQVGGVSGVSLKAHFAEPKVASD"

ORIGIN
Query Match      100.0%; Score 30; DB 6; Length 1243;

```

Best Local Similarity 100.0%; Pred. No. 0.14; Mismatches 0; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAACCGGAGGGGATCAGTACTGCGCAACA 30
|||||
Db 1005 GCAACCGGAGGGGATCAGTACTGCGCAACA 976

RESULT 15
LOCUS AE006931/c
DEFINITION Mycobacterium tuberculosis CDC1551, linear BCT 27-APR-2001
complete genome.
ACCESSION AE006931
VERSION AE000516
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 14664)
AUTHORS Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Uterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A. and Bishai, W.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES
source
1. 14664
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/notes="clinical strain"
142..552
/gene="MT0213"
142..552
/gene="MT0213"
/notes="identified by Glimmer2; putative"
/codon_start=1
/product="hypothetical protein"
/transl_table=11
/protein_id="AAK44434.1"
/db_xref="GI:13879698"
/translation="MTGATATRRLLAVLIALPGAVALLAPEASATGSDPCAAS
EVARTVGSVAKSGDYLDHPETNQVNTAVLQQVGGSGVASLKAHFEANPKVADLH
ALSQPLTDLSTRCSLPSLGLQALGMQAVQGARR"
complement (604..1842)
/gene="MT0214"
complement (604..1842)
/gene="MT0214"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44435.1"
/db_xref="GI:13879699"
/translation="NSHDAPARNLQRVGPALPRTRVGAAPPAGVPPRGKWWLRNVAL
AIVAVLAEVGLQDLAKAWVLYRAKWWMLLAVAAGAASMSFSAQIQRTLLKSA
GHVKQWRSEAFYAANSLSLTLPGGPVLSTATFLRQQRIWGAATVVASWQVMSGVL
QAVGLALLGLGGAFFLGAKNPNPSLFTLGGFTLLLLLAQAVASREPLIEGIGRVLS

gene
CDS

WANSVGRPADAGLPKWRETLMQLESVSLGRRDLGVAFKSLFNWIADVACLGAFAVA
AGDHSVGGALVAAARAVGTIFLPMFEGGLVVEAVLVPLGVSSGMPLPSAISMLLY
RLSLMLIAIGWVWFFMERFETSTADSDNDROPPTDFNLRLVLPQPGTCCDDPVEVT
PQGPATPDLRPEGGETPR
2012..3115
/gene="MT0215"
2012..3115
/gene="MT0215"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44436.1"
/db_xref="GI:13879700"
/translation="MSASLDDASVAPLVKRTAAWRFVLVLAAMVALLWAKFEVI
VPLVLAALMUSALLVPPVMDLSRGLPFAVAVALVLLSGFAVLGGIITFVVSQIAGL
PHLVEVERSDSARRWLEIEPAHLRGEQIDNAGNAIEALRNNAKLTSAUGTAAT
ITELVTAALVLFLLFLYGGRSIQWYTKAFPSVDRVRAAGRAGVSLGAYARA
TFVLVDAAGVAGLAGVMPVPLASLFFGAFIPLICAVVAGGLAVVALLAKG
IGYALITVGLLIAVNOLEHLQPLVNGRAVSIHPLAVLAIAIAGGVLAGVVGALLAV
PTVAFNNVQVLGGNPFADVADVSSDHLTEV"
complement (3112..5946)
/gene="MT0216"
complement (3112..5946)
/gene="MT0216"
/notes="similar to GP:2104606; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="membrane protein, MmpL family"
/protein_id="AAK44437.1"
/db_xref="GI:13879701"
/translation="MFAWGRGTVYRVYFVIGVMVALCLGGVGLGKHVITQSGFY
DGSQSVQVILGPDQVYGRDRSGHIVAFQAPAGKTVDVDPAMSKVVDLNRFOODHP
DQVLGAGVLRASQATGATADKKYTVFSIPLKGGDDDTILNNYKAIPDLRLDGGT
VKLAGQFVBAALGTATDORNEVIALVAVLVFVFGGVIAAGLPMVVGSLCIA
GALGMFLAIFGPVHYFAQVPSVLLGLGIAIDYGLFVSRFEEIAEGYDETAVRR
TVITAGRTVFSAVLIVASAILLFPQGLSLTYATIASVMSLSILSITVLPACLG
ILGHVDAAGVRLTFRVPLFANWKSIAAYLNWLAADRLQRTKTRREEVAGFWGLNVR
MKRPVLAAPITVIMILLIIPVGLSLGGISEKYLPTNSVRQAEFDFKLPQGYRN
PLTIVTSHQPVTDQADIADIRSKAMAGIFIPDNDPANWQERAVAVGASKDPSV
RVLQNLINPADASKLTTELRAITPPKGIITVVGGTPEALDLSIHGLFAKPLAVVIL
LTTIIVLMFAFSGVILFKATLSALTGSTMGILLIWFVDGHFSKWLNFPTPLTA
PVIILIIALVFLSTDEYFLVSRVREARGMSTQAIRITGAATGRITTAALILVA
VVGAFVPSDLMMKYLAFGLAALLLDATVRMFLVPSVMKLLGDDCWWPWARRL
QTRIGLGIHLPLDERNPVSNRGPAPVTAGLVAAAGDPDPDPDTPHPLAESPRP
ARSSPASPILTPALEATAAPASGASTTRMOIGSGSTPTTTLAAAGRSVQSPAST
PPPTTPSPASAGQTRAMPANRSTDAAGDPAETPAALPIIRSDGDDSEAAETEQLN
ARGISDKTRQRRRGGLSAQDLLRREGRL"
complement (6012..6746)
/gene="MT0217"
complement (6012..6746)
/gene="MT0217"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44438.1"
/db_xref="GI:13879702"
/translation="MMSLRTDVTQSSELSARHSLAEDLSQDLTSLGAPGARVLL
VDAENLDMGLGSILGRRTALRPRDGLRLLARTAEIVAGRPGISTEPETVFT
NIAPGSREVTPWVALENTGVFAFKVDESDVDRLMLAIDERVREGLAALVVA
SADQAFRQPLEAVARSGTPVQLGFREHASWALASDTLEFVDIEDIAGVREPLPRI
GLDSLPEQAGLQFPRPLUSLLTSRV"
complement (6743..7534)
/gene="MT0218"
complement (6743..7534)
/gene="MT0218"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK44439.1"

Search completed: April 29, 2004, 06:01:33
Job time : 455.006 secs

/db_xref="GI:13879703"
/translation="MVHHGQMAQPGVGLRPDTPVAGQLPSTSISSRSRSGISKAQRE
TWRLWPELGLLALPOSPRTPTVDTRAFGRDAPVLEIGSGSGTSTLAWAKAEPHVD
VIADVVRGLGALLCAIDVSGSDGILNIIILGNVNDVLQHLIADPLSLGCVAFVFPDP
WPKARHKKRLLQAPATWALLADRLVPSGVLAATDHPGYAEHIAAAGADAEPLRVDP
DTELLPISVVRPATKYERKAQLGGGAVIELLKKGGCSEKDLKIR"

gene
CDS

7666..8751
/gene="MT0219"
7666..8751
/gene="MT0219"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44440.1"
/db_xref="GI:13879704"

/translation="MSGQGHQIFVDELARFATSSADRVVAIAQRAAEPLRVAVGRP
GVGCRVARALQAGSSGSMVTVPQARAADSDVDLVVYTVVEVVKEDREAIANRRP
VVAIVNKADLAGLPGAGPIVMAQCAQFSTLLGVPMESMIGLLAAALDDDDTLR
AVLRALAAHDFDADRAVAGLAAALPVTEVRLRLDLDLFGALGMAAFRPG
PSRTPAQLRTLLRRVSGVDVADIKVTAAGSEVEYRRLDAVAEALAAQAKEIGGPI
GEFLRDDTVLARMAAADVALAVGLDVGPLDDPAAHLPRAVRWHRYSLDNGDMRTG
GADIARGLRLWLSLGGMLHRYKSS"

gene
CDS

8781..10226
/gene="MT0220"
8781..10226
/gene="MT0220"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44441.1"
/db_xref="GI:13879705"

/translation="MDLVAAIAIPAGLAGLPLVINRREVVLVTGPMACVSGVRAALA
ERLPQRFLVTAELGPDAPVAVFVVSATALTSDCVLLDTAAEDTDAVAVVSKI
DVHGRVDFVTSNRDLARARVAVFVWVAAAAPGLGEPYLDLVAAIQKQADPA
VARNMLRAVESLLVARRFDGDAOSAGRRARVDALROCRVTLRQGRCSKSEHTIA
LRAQIOHARVLSFYARNPCSLRLVELOEVLGSLKDIARFAATRGVQEVAVG
EGAVAHADVAQLGLGVQPVULENPVAVLPVAPPLISRLRIKTLITLLGAGFLG
IAUTLRLVAGLTPGAAGMVAIGLAVTAWVNAALLHDRVVDVDRVTGEVTAS
LRSVVQLVATRVVATLLSTAISSERDDAENARVADQVSIIDGELLEHAFAARA
LRDEMPAVRAALEAVRAELGEPGPTGLF"

gene
CDS

10410..12230
/gene="MT0221"
10410..12230
/gene="MT0221"
/note="similar to PID:2104609; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate carboxykinase"
/protein_id="AAK44442.1"
/db_xref="GI:13879706"

/translation="WTSATIPGLDTPATHQGLLSWEEVVELTQDPRVFTDGSSEE
FQRLCPQVLAAGTFLINPEKKNYSALSDPSDVARVESRTYICSAKEIDAGPTNNW
MDPEGRVSKMDLYRCMGRTWYVVPFCVGLGADPKLGEITDSEYVVVWSMTMT
RMGKAALKMGDDGPFVKALHSVGALEPCQKQVAMPSETKYITHFPETREIWSYGS
GYGENALLGKCYSLASIAAMAHDEGLAEHMLILKLISPENKAVYFAAAPPSACGKT
NLAMLQPTIPGRVETLDDDIAMWRFGDGRLYAVNPERGFEVAPGTNNKSNPMR
TIAAGNTVFNVAITDDGDDWEGLEGDFQHLIDMKNWDWYFRETETNAAHNSRYCT
PMSCQPIAPEDPDQVPSIGLFGGRRTVPLVTEARDWQGVFIGATLGSQTA
AMEGKRVNVRDPMALPFLGVNGVDYFQHWNLGKHADSKLPKVFYVWFRGDDG
RFLWPGFGENSRLVWIDRIEIKAGGATPIGTVPVAVEDLDLGLDGLDADVAALA
VDADENRQELPLEEWLQVFVEKLPVGVKDFDALKERLG"

Query Match 100.0%; Score 30; DB 1; Length 14664;
Best Local Similarity 100.0%; Pred. No. 0.089;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAACGCGAGGCGGATCAGTACTGCCAACA 30
DB 205 GCAACGCGAGGCGGATCAGTACTGCCAACA 176

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:01:30 ; Search time 2365.53 Seconds

(without alignments)
378.717 Million cell updates/sec

Title: US-10-624-714-4

Perfect score: 3
Sequence: 1 gcaacgcaggcgatcagctactgcaaca 30

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_estc:*

9: gb_est1:*

10: gb_est2:*

11: gb_est3:*

12: gb_est4:*

13: gb_est5:*

14: gb_est6:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pbg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	21	70.0	1590	13	BU434548
2	20.6	68.7	501	28	BH650137
3	20.6	68.7	513	12	BG587118
4	20.6	68.7	558	28	BH663254

C 5	20.6	68.7	614	28	BH458651
C 6	20.6	68.7	619	28	BH699403
C 7	20.6	68.7	658	28	BZ088759
8	20.6	68.7	677	28	BH495798
9	20.6	68.7	679	28	BH553789
10	20.6	68.7	687	28	BH570958
11	20.6	68.7	690	28	BH432840
C 12	20.6	68.7	696	28	BH449191
C 13	20.6	68.7	696	28	BH707351
C 14	20.6	68.7	711	28	BH737846
C 15	20.6	68.7	715	28	BH484743
C 16	20.6	68.7	738	28	BZ519113
C 17	20.6	68.7	740	28	BZ513022
C 18	20.6	68.7	752	28	BZ427742
C 19	20.6	68.7	758	28	BZ071565
C 20	20.6	68.7	782	28	BH671560
C 21	20.6	68.7	794	28	BZ457050
C 22	20.6	68.7	795	28	BH549269
C 23	20.6	68.7	800	28	BZ482950
C 24	20.6	68.7	803	28	BH558234
C 25	20.6	68.7	811	28	BZ451468
C 26	20.6	68.7	826	28	BH442819
C 27	20.6	68.7	830	28	BZ498232
C 28	20.6	68.7	833	28	BH651621
C 29	20.6	68.7	837	28	BZ434938
C 30	20.6	68.7	839	28	BH645109
C 31	20.6	68.7	862	28	BZ443065
C 32	20.6	68.7	869	28	BH685443
C 33	20.6	68.7	878	28	BZ454497
C 34	20.6	68.7	939	14	CB686143
C 35	20.6	68.7	1027	28	BZ433580
C 36	20.6	68.7	1103	28	BZ500089
C 37	20.4	68.0	466	9	AV960672
C 38	20.4	68.0	557	12	BG366351
C 39	20.4	68.0	642	13	BQ469880
C 40	20.4	68.0	903	10	BE570639
C 41	20	66.7	689	13	BX305062
C 42	20	66.7	691	29	CNS02UOF
C 43	20	66.7	966	29	CNS02B75
C 44	19.6	65.3	694	28	CC058604
C 45	19.6	65.3	799	28	BH568716

ALIGNMENTS

RESULT 1
BU434548
LOCUS 603220134F1 CSEQRBN10 Gallus gallus CDNA clone CHEST213c9 5', mRNA
DEFINITION sequence.
ACCESSION BU434548
VERSION BU434548.1 GI:25923859
KEYWORDS EST
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1590)
BOARDMAN, P.E., SANZ-EZQUERRO, J., OVERTON, I.M., BURT, D.W., BOSCH, E., FONG, W.T., TICKLE, C., BROWN, W.R.A., WILSON, S.A. and HUBBARD, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392
COMMENT Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409

BU434548 1590 bp mRNA linear EST 29-NOV-2002
603220134F1 CSEQRBN10 Gallus gallus CDNA clone CHEST213c9 5', mRNA
sequence.
BU434548
GI:25923859
Gallus gallus (chicken)
Gallus gallus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 1590)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source
1. .1590
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Layer and broiler"
/db_xref="taxon:9031"
/clone="CHEST213c9"
/sex="Male and female"
/tissue_type="Chondrocytes isolated from growth plate cartilage"
/dev_stage="adult"
/lab_hosts="DH10B"
/clone_lib="CSEORBN10"
/note="Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Ronaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN
Query Match 70.0%; Score 21; DB 13; Length 1590;
Best Local Similarity 82.8%; Pred. No. 1.1e+03;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAACGCGAGCGGCGATCAGTACTGCCAACA 30
|||||
DB 906 CAAACGAGGCGCATCAATCTGCCAACA 934

RESULT 2
BH650137 501 bp DNA linear GSS 19-FEB-2002
LOCUS BOMGC22TR BO_2_3_KB Brassica oleracea genomic clone BOMGC22,
DEFINITION genomic survey sequence.
ACCESSION BH650137 GI:18708022
VERSION BH650137.1
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1. (bases 1 to 501)
TOWN.C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
UNPUBLISHED (2001)
CONTACT: Chris Town
TIGR
7912 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TR
Class: sheared ends.

FEATURES
source
1. .501
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="TO1000DH3"
/db_xref="taxon:3712"
/clone="BOMGC22"
/clone_lib="BO_2_3_KB"

ORIGIN

Query Match 68.7%; Score 20.6; DB 28; Length 501;
Best Local Similarity 85.2%; Pred. No. 1.2e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCACGCGAGCGGCGATCAGTACTGCCA 27
|||||
DB 470 CCAAAGCTAGGCGGTTCAGTACGGCCA 496

RESULT 3

BG587118
LOCUS BG587118 513 bp mRNA linear EST 11-APR-2001
DEFINITION EST488888 MHAM Medicago truncatula/Glomus versiforme mixed EST library cDNA clone pMHAM-48E3 5' end, mRNA sequence.
ACCESSION BG587118 GI:13602182
VERSION BG587118.1
KEYWORDS EST.
SOURCE Medicago truncatula/Glomus versiforme mixed EST library
ORGANISM Medicago truncatula/Glomus versiforme mixed EST library
Eukaryota; mixed EST libraries.
REFERENCE 1. (bases 1 to 513)
AUTHORS Harrison,M.J., Liu,J., Town,C.D., Van Aken,S., Utterback,T., Cho,J. and Fraser,C.M.
ESTs from roots of Medicago truncatula after colonization with Glomus versiforme, 2001
UNPUBLISHED (2001)
CONTACT: Harrison M.J.
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
Noble EST name: N386186 TIGR sequence name: MTDDC26TK More
information is available at: <http://www.medicago.org>
Seq primer: SKmod (CTA GAA CTA gtg gat CC).

FEATURES

Location/Qualifiers
1. .513
/organism="Medicago truncatula/Glomus versiforme mixed EST library"
/mol_type="mRNA"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="pMHAM-48E3"
/tissue_type="roots colonized with Glomus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages."
/lab_hosts="E. coli strain XL0LR"
/clone_lib="MHAM"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the Unizap XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-assist helper phage and propagated in XL0LR cells."

ORIGIN

Query Match 68.7%; Score 20.6; DB 12; Length 513;
Best Local Similarity 85.2%; Pred. No. 1.2e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCACGCGAGCGGCGATCAGTACTGCCA 27
|||||
DB 63 CCAAAGCTAGGCGGTTCAGTACGGCCA 89

```

RESULT 4
BH663254/c
LOCUS      BH663254      558 bp      DNA      linear      GSS 19-FEB-2002
DEFINITION BOMJR04TF BO_2_3_KB Brassica oleracea genomic clone BOMJR04,
            genomic survey sequence.
ACCESSION  BH663254
VERSION     BH663254.1 GI:118722170
KEYWORDS   GSS.
SOURCE      Brassica oleracea
ORGANISM    Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 558)
AUTHORS     Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
COMMENT     Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
    1..558
       /organism="Brassica oleracea"
       /mol_type="genomic DNA"
       /strain="TO1000DH3"
       /db_xref="taxon:3712"
       /clone="BOMJR04"
       /clone_lib="BO_2_3_KB"
       /note="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared
       genomic DNA inserted into pHOSt using BstXI linkers"

FEATURES             source
    source
    Seq primer: TF
    Class: sheared ends.
    Location/Qualifiers
        1..558
           /organism="Brassica oleracea"
           /mol_type="genomic DNA"
           /strain="TO1000DH3"
           /db_xref="taxon:3712"
           /clone="BOMJR04"
           /clone_lib="BO_2_3_KB"
           /note="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared
           genomic DNA inserted into pHOSt using BstXI linkers"

ORIGIN
Query Match      68.7%; Score 20.6; DB 28; Length 558;
Best Local Similarity 85.2%; Pred. No. 1.2e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCAACGCGAGGGCGATCAGTACTGCCA 27
    |||||
Db 139 GCAAGCTAGGCGGCTTCAGTACGGCCA 113

RESULT 5
BH458651/c
LOCUS      BH458651      614 bp      DNA      linear      GSS 13-DEC-2001
DEFINITION BOGQC38TR BOGQ Brassica oleracea genomic clone BOGQC38, genomic
            survey sequence.
ACCESSION  BH458651
VERSION     BH458651.1 GI:117648396
KEYWORDS   GSS.
SOURCE      Brassica oleracea
ORGANISM    Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 614)
AUTHORS     Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
COMMENT     Other GSSs: BOGQC38TF
            Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
    1..619
       /organism="Brassica oleracea"
       /mol_type="genomic DNA"
       /strain="TO1000DH3"
       /db_xref="taxon:3712"
       /clone="BOHVO28"
       /clone_lib="BO_2_3_KB"
       /note="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared
       genomic DNA inserted into pHOSt using BstXI linkers"

FEATURES             source
    source
    Seq primer: TF
    Class: sheared ends.
    Location/Qualifiers
        1..619
           /organism="Brassica oleracea"
           /mol_type="genomic DNA"
           /strain="TO1000DH3"
           /db_xref="taxon:3712"
           /clone="BOHVO28"
           /clone_lib="BO_2_3_KB"
           /note="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared
           genomic DNA inserted into pHOSt using BstXI linkers"

ORIGIN
Query Match      68.7%; Score 20.6; DB 28; Length 619;
Best Local Similarity 85.2%; Pred. No. 1.2e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCAACGCGAGGGCGATCAGTACTGCCA 27
    |||||
Db 507 GCAAGCTAGGCGGCTTCAGTACGGCCA 481

RESULT 6
BH699403/c
LOCUS      BH699403      619 bp      DNA      linear      GSS 20-FEB-2002
DEFINITION BOHVO28TF BO_2_3_KB Brassica oleracea genomic clone BOHVO28,
            genomic survey sequence.
ACCESSION  BH699403
VERSION     BH699403.1 GI:18774647
KEYWORDS   GSS.
SOURCE      Brassica oleracea
ORGANISM    Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1 (bases 1 to 619)
AUTHORS     Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
TITLE       Whole genome shotgun sequencing of Brassica oleracea
JOURNAL     Unpublished (2001)
COMMENT     Other GSSs: BOHVO28TR
            Contact: Chris Town

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.
Location/Qualifiers
    1..619
       /organism="Brassica oleracea"
       /mol_type="genomic DNA"
       /strain="TO1000DH3"
       /db_xref="taxon:3712"
       /clone="BOHVO28"
       /clone_lib="BO_2_3_KB"
       /note="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared
       genomic DNA inserted into pHOSt using BstXI linkers"

FEATURES             source
    source
    Seq primer: TF
    Class: sheared ends.
    Location/Qualifiers
        1..619
           /organism="Brassica oleracea"
           /mol_type="genomic DNA"
           /strain="TO1000DH3"
           /db_xref="taxon:3712"
           /clone="BOHVO28"
           /clone_lib="BO_2_3_KB"
           /note="Vector: pHOSt; Site 1: BstXI; 2-3 kb sheared
           genomic DNA inserted into pHOSt using BstXI linkers"

ORIGIN
Query Match      68.7%; Score 20.6; DB 28; Length 619;
Best Local Similarity 85.2%; Pred. No. 1.2e+03;
Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCAACGCGAGGGCGATCAGTACTGCCA 27
    |||||
Db 263 GCAAGCTAGGCGGCTTCAGTACGGCCA 237

RESULT 7
B2088759/c

```

LOCUS BZ088759 658 bp DNA linear GSS 10-OCT-2002
 DEFINITION lxf93f03.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
 sequence.
 ACCESSION BZ088759
 VERSION BZ088759.1 GI:23723077
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE 1 (bases 1 to 658)
 AUTHORS Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,
 Nash, W., Rabinowicz, P.D. and Wilson, R.K.
 TITLE Whole genome shotgun reads from Brassica oleracea
 JOURNAL Unpublished (2002)
 COMMENT Contact: Richard K. Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@watson.wustl.edu
 Plate: lxf93 row: f column: 03
 Seq primer: -28RppOT reverse
 Class: shotgun
 High quality sequence start: 8
 High quality sequence stop: 543.
 FEATURES
 source
 1.658
 Location/Qualifiers
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /db_xref="taxon:3712"
 /clone_lib="B.oleracea002"
 /note="vector: pOTw13; whole genome shotgun library from
 flowering buds. DNA was purified from a crude nuclear
 prep using Brassica oleracea TO1000DH3 buds provided by
 Thomas Osborn at the University of Wisconsin. Genomic
 DNA was provided by Pablo Rabinowicz (CSHL) and the
 shotgun library prepared at Washington University Genome
 Sequencing Center."
 ORIGIN
 Query Match 68.7%; Score 20.6; DB 28; Length 658;
 Best Local Similarity 85.2%; Pred. No. 1.3e+03;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GCACGCGAGGGCGATCAGTACTGCCA 27
 DB 463 GCACGCTAGGGCGGTTTCAGTACGCCA 437
 RESULT 8
 BZ495798
 LOCUS BZ495798 677 bp DNA linear GSS 13-DEC-2001
 DEFINITION BGSB39TR BGS Brassica oleracea genomic clone BGSB39, genomic
 survey sequence.
 ACCESSION BZ495798
 VERSION BZ495798.1 GI:17703902
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE 1 (bases 1 to 677)
 AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
 TITLE Whole genome shotgun sequencing of Brassica oleracea
 JOURNAL Unpublished (2001)
 COMMENT Other GSSs: BGSB39TF
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TR
 Class: sheared ends.
 FEATURES
 source
 1..677
 Location/Qualifiers
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BGSB39"
 /clone_lib="BGS"
 /note="Vector: pBOS1; Site 1: BstXI; 2-3 kb sheared
 genomic DNA inserted into pBOS1 using BstXI linkers"
 ORIGIN
 Query Match 68.7%; Score 20.6; DB 28; Length 677;
 Best Local Similarity 85.2%; Pred. No. 1.3e+03;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GCACGCGAGGGCGATCAGTACTGCCA 27
 DB 71 GCACGCTAGGGCGGTTTCAGTACGCCA 97
 RESULT 9
 BZ953789
 LOCUS BZ953789 679 bp DNA linear GSS 01-OCT-2002
 DEFINITION odi85d07.g1 B.oleracea002 Brassica oleracea genomic, genomic survey
 sequence.
 ACCESSION BZ953789
 VERSION BZ953789.1 GI:23435016
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE 1 (bases 1 to 679)
 AUTHORS Delehaunty, K., Fewell, G., Fulton, L., McCombie, W.R., Miner, T.,
 Nash, W., Rabinowicz, P.D. and Wilson, R.K.
 TITLE Whole genome shotgun reads from Brassica oleracea
 JOURNAL Unpublished (2002)
 COMMENT Contact: Richard K. Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@watson.wustl.edu
 Plate: odi85 row: d column: 07
 Seq primer: -28RppOT reverse
 Class: shotgun
 High quality sequence start: 47
 High quality sequence stop: 512.
 FEATURES
 source
 1.679
 Location/Qualifiers
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /db_xref="taxon:3712"
 /clone_lib="B.oleracea002"
 /note="Vector: pOTw13; whole genome shotgun library from
 flowering buds. DNA was purified from a crude nuclear
 prep using Brassica oleracea TO1000DH3 buds provided by
 Thomas Osborn at the University of Wisconsin. Genomic
 DNA was provided by Pablo Rabinowicz (CSHL) and the
 shotgun library prepared at Washington University Genome
 Sequencing Center."
 ORIGIN
 Query Match 68.7%; Score 20.6; DB 28; Length 679;
 Best Local Similarity 85.2%; Pred. No. 1.3e+03;
 Matches 23; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 GCACGCGAGGGCGATCAGTACTGCCA 27
 DB 576 GCACGCTAGGGCGGTTTCAGTACGCCA 602

RESULT 10	BH570958	687 bp	DNA	linear	GSS 14-DEC-2001
LOCUS	BOHF34TF BOHF Brassica oleracea genomic clone BOHF34, genomic				
DEFINITION	survey sequence.				
ACCESSION	BH570958				
VERSION	BH570958.1	GI:17822797			
KEYWORDS	GSS.				
SOURCE	Brassica oleracea				
ORGANISM	Brassica oleracea				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.				
AUTHORS	1 (bases 1 to 687)				
TITLE	Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.				
JOURNAL	Whole genome shotgun sequencing of Brassica oleracea				
COMMENT	Unpublished (2001)				
	Other GSSs: BOHF34TR				
	Contact: Chris Town				
FEATURES	TIGR				
source	9712 Medical Center Drive, Rockville, MD 20850, USA.				
	tel: 301-838-3523				
	Fax: 301-838-0208				
	Email: cdtown@tigr.org				
	DNA is from a doubled haploid provided by Tom Osborn.				
	Seq primer: TF				
	Class: sheared ends.				
	Location/Qualifiers				
	1..687				
	/organism="Brassica oleracea"				
	/mol_type="genomic DNA"				
	/strain="TO1000DH3"				
	/db_xref="taxon:3712"				
	/clone="BOHF34"				
	/clone_lib="BOHF"				
	/note="vector: PHOS1; Site_1: BstXI; 2-3 kb sheared				
	genomic DNA inserted into PHOS1 using BstXI linkers"				
ORIGIN					
	Query Match	68.7%	Score 20.6;	DB 28;	Length 687;
	Best Local Similarity	85.2%;	Pred.No. 1.3e+03;		
	Matches	23;	Conservative	0;	Mismatches 4;
				Indels	0;
				Gaps	0;
QY	1	GCAACGCGAGCGGCGATCAGTACTGCCA	27		
DB	314	GCAAGCTAGGCGGCTTCAGTACGCCA	340		
RESULT 11	BH432840	690 bp	DNA	linear	GSS 12-DEC-2001
LOCUS	BOGX81TF BOGX Brassica oleracea genomic clone BOGX81, genomic				
DEFINITION	survey sequence.				
ACCESSION	BH432840				
VERSION	BH432840.1	GI:17618561			
KEYWORDS	GSS.				
SOURCE	Brassica oleracea				
ORGANISM	Brassica oleracea				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.				
AUTHORS	1 (bases 1 to 690)				
TITLE	Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.				
JOURNAL	Whole genome shotgun sequencing of Brassica oleracea				
COMMENT	Unpublished (2001)				
	Other GSSs: BOGX81TR				
	Contact: Chris Town				
FEATURES	TIGR				
source	9712 Medical Center Drive, Rockville, MD 20850, USA.				
	tel: 301-838-3523				
	Fax: 301-838-0208				

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 06:08:26 ; Search time 560 Seconds
(without alignments)
242.328 Million cell updates/sec

Title: US-10-624-714-4

Perfect score: 30
Sequence: 1 gcaacgagggcgatcagctactgccaaca 30

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

```
1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/prodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19.6	65.3	1428	15	US-10-156-761-3841
C 2	19.6	65.3	9025608	15	Sequence 3841, Ap
C 3	19.4	64.7	1125	16	US-10-156-761-1
C 4	19.4	64.7	1128	16	Sequence 39546, A
C 5	19.4	64.7	32050	13	US-10-369-493-39174
C 6	18.6	62.0	1176	13	US-10-087-192-793
C 7	18.6	62.0	38734	10	Sequence 793, App
C 8	18.6	62.0	38734	11	US-10-424-599-97077
C 9	18.4	61.3	447	13	Sequence 30, Appl
C 10	18.4	61.3	735	13	US-09-373-658-30
C 11	18.4	61.3	809	13	Sequence 30, Appl
C 12	18.4	61.3	816	13	US-10-282-122A-23074
C 13	18.4	61.3	864	15	Sequence 24130, A
C 14	18.4	61.3	1311	13	Sequence 1392, Ap
					US-10-425-114-1392
					Sequence 32478, A
					Sequence 3578, Ap
					Sequence 17861, A

```

C 15 18.4 61.3 1472 13 US-10-425-114-24079
C 16 18.2 60.7 230 15 US-10-029-386-17173
C 17 18.2 60.7 593 15 US-10-029-386-3473
C 18 18.2 60.7 1254 15 US-10-156-761-5096
C 19 17.8 59.3 540 13 US-10-282-122A-23020
C 20 17.8 59.3 720 13 US-10-027-632-136213
C 21 17.8 59.3 720 13 US-10-027-632-136214
C 22 17.8 59.3 720 16 US-10-027-632-136214
C 23 17.8 59.3 720 16 US-10-027-632-136214
C 24 17.8 59.3 876 16 US-10-429-802-19
C 25 17.8 59.3 876 16 US-10-430-503-10
C 26 17.8 59.3 1215 16 US-10-369-493-44376
C 27 17.8 59.3 1231 9 US-09-974-300-309
C 28 17.8 59.3 1341 13 US-10-282-122A-24060
C 29 17.8 59.3 1380 13 US-10-282-122A-23964
C 30 17.8 59.3 2334 15 US-10-156-761-3217
C 31 17.8 59.3 2679 15 US-10-156-761-176
C 32 17.8 59.3 3249 9 US-09-738-626-3414
C 33 17.8 59.3 4265 13 US-10-058-024-10
C 34 17.8 59.3 42115 17 US-10-432-422-59
C 35 17.8 59.3 68750 14 US-10-014-717-1
C 36 17.8 59.3 744802 16 US-10-292-798-1369
C 37 17.8 59.3 3309400 9 US-09-738-626-1
C 38 17.8 59.3 9025608 15 US-10-156-761-1
C 39 17.6 58.7 1023 13 US-10-282-122A-28792
C 40 17.6 58.7 1039 13 US-10-282-122A-28795
C 41 17.4 58.0 284 13 US-10-282-122A-13499
C 42 17.4 58.0 475 13 US-10-027-632-276905
C 43 17.4 58.0 475 16 US-10-027-632-276905
C 44 17.4 58.0 681 13 US-10-027-632-110430
C 45 17.4 58.0 681 13 US-10-027-632-110431

```

ALIGNMENTS

RESULT 1

```

US-10-156-761-3841/c
; Sequence 3841, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HOSHIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3841
; LENGTH: 1428
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1428)
US-10-156-761-3841

```

Query Match 65.3%; Score 19.6; DB 15; Length 1428;

Best Local Similarity 84.6%; Pred. No. 22;

Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGCGAGGGCGATCAGTACTGCCA 27

Db 1383 CAACGCGAGGGCGATCAGTACTGCCA 1358

```
RESULT 2
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 65.3%; Score 19.6; DB 15; Length 9025608;
Best Local Similarity 84.6%; Pred. No. 53;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CAACGCGAGCGGCATCAGTACTGCCA 27
DB 4764072 CAACGCGCGGCATCCTGCTACCGCCA 4764097

RESULT 3
US-10-369-493-39546
; Sequence 39546, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 39546
; LENGTH: 1125
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-10-369-493-39546

Query Match 64.7%; Score 19.4; DB 16; Length 1125;
Best Local Similarity 95.2%; Pred. No. 27;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 GCGAGGCGGCATCAGTACTGCC 26
DB 950 GCGAGGCGGCATCTGTACTGCC 970

RESULT 4
US-10-369-493-39174
; Sequence 39174, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 39174
; LENGTH: 1128
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
US-10-369-493-39174

Query Match 64.7%; Score 19.4; DB 16; Length 1128;
Best Local Similarity 95.2%; Pred. No. 27;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 6 GCGAGGCGGCATCAGTACTGCC 26
DB 950 GCGAGGCGGCATCTGTACTGCC 970

RESULT 5
US-10-087-192-793
; Sequence 793, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 793
; LENGTH: 32050
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(32050)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-793

Query Match 64.7%; Score 19.4; DB 13; Length 32050;
Best Local Similarity 79.3%; Pred. No. 38;
Matches 23; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CAACGCGAGCGGCATCAGTACTGCCA 30
DB 18041 CCAAGCCAGCGCGCTGCTGCTGCCCA 18069

RESULT 6
```

```
US-10-424-599-97077/c
; Sequence 97077, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 97077
; LENGTH: 1176
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_58672C.1
US-10-424-599-97077

Query Match          62.0%; Score 18.6; DB 13; Length 1176;
Best Local Similarity 84.0%; Pred. No. 65;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCAACGGCGGGGATCAGTACTGTC 25
DB 114 GCAACGGAGTGGGAACAGTACTGC 90

RESULT 7
US-09-373-658-30
; Sequence 30, Application US/09373658
; Publication No. US20030092900A1
; GENERAL INFORMATION:
; APPLICANT: Iruela-Arispe, Luisa
; APPLICANT: Hastings, Gregg A.
; APPLICANT: Ruben, Steven M.
; APPLICANT: Jonak, Zdenka L.
; APPLICANT: Trullis, Stephen H.
; APPLICANT: Fronwald, James A.
; APPLICANT: Terrett, Jonathan A.
; TITLE OF INVENTION: Meth1 and Meth2 Polynucleotides and Polypeptides
; FILE REFERENCE: 1488.1070006
; CURRENT APPLICATION NUMBER: US/09/373,658
; CURRENT FILING DATE: 1999-08-13
; NUMBER OF SEQ ID NOS: 125
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 38734
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:Unknown
US-09-373-658-30

Query Match          62.0%; Score 18.6; DB 10; Length 38734;
Best Local Similarity 84.0%; Pred. No. 94;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 CGCAGGGCGGATCAGTACTGCCAAC 29
DB 29424 CACGAGGCGATCCGTCCTGCCAGC 29448

RESULT 8
US-09-989-687-30
; Sequence 30, Application US/09989687
; Publication No. US20040002449A1
; GENERAL INFORMATION:
; APPLICANT: Hastings, Gregg A.
; APPLICANT: Ruben, Steven M.
```

```
US-10-424-599-97077/c
; TITLE OF INVENTION: Meth1 and Meth2 Polynucleotides and Polypeptides
; FILE REFERENCE: 1488.107000D
; CURRENT APPLICATION NUMBER: US/09/989,687
; CURRENT FILING DATE: 2001-11-21
; NUMBER OF SEQ ID NOS: 126
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 38734
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism:Unknown
US-09-989-687-30

Query Match          62.0%; Score 18.6; DB 11; Length 38734;
Best Local Similarity 84.0%; Pred. No. 94;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 CGCAGGGCGGATCAGTACTGCCAAC 29
DB 29424 CACGAGGCGATCCGTCCTGCCAGC 29448

RESULT 9
US-10-282-122A-23074/c
; Sequence 23074, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23074
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-23074
```

```
Query Match          61.3%; Score 18.4; DB 13; Length 447;
Best Local Similarity 78.6%; Pred. No. 73;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CAACGCGAGGCGGATCAGTACTGCGCAAC 29
Db 369 CAACGCGAGGCGGATGAGCGCGGCAGC 342

RESULT 10
US-10-282-122A-24130/c
; Sequence 24130, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/131,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24130
; LENGTH: 735
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-10-282-122A-24130

Query Match          61.3%; Score 18.4; DB 13; Length 735;
Best Local Similarity 78.6%; Pred. No. 77;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CAACGCGAGGCGGATCAGTACTGCGCAAC 29
Db 369 CAACGCGAGGCGGATGAGCGCGGCAGC 342

RESULT 11
US-10-425-114-1392
; Sequence 1392, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 1392
; LENGTH: 809
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700152006_FLI
US-10-425-114-1392

Query Match          61.3%; Score 18.4; DB 13; Length 809;
Best Local Similarity 78.6%; Pred. No. 78;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CAACGCGAGGCGGATCAGTACTGCGCAAC 29
Db 339 CAACGCGAGGCGGATGAGCGCGGCAC 366

RESULT 12
US-10-425-114-32478
; Sequence 32478, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 32478
; LENGTH: 816
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73376C06_FLI
US-10-425-114-32478

Query Match          61.3%; Score 18.4; DB 13; Length 816;
Best Local Similarity 78.6%; Pred. No. 78;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 CAACGCGAGGCGGATCAGTACTGCGCAAC 29
Db 339 CAACGCGAGGCGGATGAGCGCGGCAC 366

RESULT 13
US-10-156-761-3578/c
; Sequence 3578, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
```

```

Search completed: April 29, 2004, 20:44:06
Job time : 578 secs

/ CURRENT REFERENCE NUMBER: 03/10/2024,1224
/ PRIOR FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:23:09 ; Search time 247.647 Seconds
(without alignments)
514.627 Million cell updates/sec.

Title: US-10-624-714-4
Perfect score: 30
Sequence: 1 gcaacgcgagggcagtcactgccaaca 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N Geneseq_29Jan04:*
- 1: Geneseqn1980s:*
 - 2: Geneseqn1990s:*
 - 3: Geneseqn2000s:*
 - 4: Geneseqn2001as:*
 - 5: Geneseqn2001bs:*
 - 6: Geneseqn2002s:*
 - 7: Geneseqn2003as:*
 - 8: Geneseqn2003bs:*
 - 9: Geneseqn2003cs:*
 - 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	30	100.0	411	2 AAX34005	Aax34005 Mycobacte
C 2	30	100.0	411	2 AAC97021	Aac97021 Mycobacte
C 3	30	100.0	498	2 AAX34006	Aax34006 Mycobacte
C 4	30	100.0	959	2 AAX34004	Aax34004 Mycobacte
C 5	30	100.0	960	2 AAX34003	Aax34003 Mycobacte
C 6	30	100.0	961	2 AAX34002	Aax34002 Mycobacte
C 7	30	100.0	1063	2 AAX34007	Aax34007 Mycobacte
C 8	30	100.0	1243	2 AAX34001	Aax34001 Mycobacte
C 9	30	100.0	110000	4 AA199682_02	Continuation (3 of
C 10	30	100.0	110000	4 AA199683_02	Continuation (3 of
C 11	19.4	64.7	110000	2 AA201425_03	Continuation (4 of
C 12	19	63.3	3328	4 AAD03951	Aad03951 A. nidula
C 13	19	63.3	3328	4 AAD03952	Aad03952 Alternati
C 14	18.8	62.7	1163	4 ABU28961	Abu28961 Drosophil
C 15	18.8	62.7	1166	4 ABU28969	Abu28969 Drosophil
C 16	18.8	62.7	2404	4 ABU23108	Abu23108 Drosophil
C 17	18.8	62.7	2936	4 ABU11577	Abu11577 Drosophil
C 18	18.8	62.7	5998	4 ABU11576	Abu11576 Drosophil
C 19	18.8	62.7	6779	4 ABU28968	Abu28968 Drosophil
C 20	18.8	62.7	6779	4 ABU28960	Abu28960 Drosophil
C 21	18.6	62.0	684	2 AAV23487	Aav23487 Pseudomon
C 22	18.6	62.0	684	3 AAA13901	Aaa13901 Pseudomon
C 23	18.6	62.0	684	6 AAD22878	Aad22878 Pseudomon

C 24	18.6	62.0	17612	2 AAV23494	Aav23494 Pseudomon
C 25	18.6	62.0	17612	3 AAA13905	Aaa13905 Pseudomon
C 26	18.6	62.0	17612	4 AAF30870	Aaf30870 Pseudomon
C 27	18.6	62.0	17612	6 AAD22882	Aad22882 Pseudomon
C 28	18.6	62.0	38734	2 AAZ32020	Aaz32020 Human MET
C 29	18.6	62.0	38734	5 AAC90077	Aac90077 Human MET
C 30	18.4	61.3	318	6 ABQ90397	Abq90397 M. capsul
C 31	18.4	61.3	447	7 ACA35204	Aca35204 Prokaryot
C 32	18.4	61.3	735	7 ACA36260	Aca36260 Prokaryot
C 33	18.4	61.3	1311	7 ACA29991	Aca29991 Prokaryot
C 34	18.4	61.3	1539	3 AAC34892	Aac34892 Arabidops
C 35	18.4	61.3	3772	4 ABL02990	Abi02990 Drosophil
C 36	18.2	60.7	1050	4 ABL21369	Abi21369 Drosophil
C 37	18.2	60.7	2899	4 ABL16150	Abi16150 Drosophil
C 38	18.2	60.7	3384	4 ABL21368	Abi21368 Drosophil
C 39	18	60.0	527	4 AAH13215	Aah13215 Human CDN
C 40	18	60.0	2231	5 ABA16032	Abal6032 Human ner
C 41	18	60.0	4078	4 AAH16633	Aah16633 Human CDN
C 42	18	60.0	4715	5 ABA16035	Abal6035 Human ner
C 43	18	60.0	4715	5 ABA16034	Abal6034 Human ner
C 44	17.8	59.3	409	6 ABN73624	Abn73624 Bovine em
C 45	17.8	59.3	539	4 AAH09943	Aah09943 Human CDN

ALIGNMENTS

RESULT 1
AAX34005/c
ID AAX34005 standard; DNA; 411 BP.
XX
AC AAX34005;
XX
DT 06-JUL-1999 (first entry)
DE Mycobacterium species nucleic acid sequence 1D.
XX
KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection; ss.
XX
OS Mycobacterium sp.
XX
FN WO9909186-A2.
XX
PD 25-FEB-1999.
XX
PF 14-AUG-1998; 98WO-FR001813.
XX
PR 14-AUG-1997; 97FR-00010404.
PR 11-SEP-1997; 97FR-00011325.
XX
(INSP) INST PASTEUR.
Gicquel B, Portnoie D, Lim B, Pelicic V, Guigueno A;
Goguet De La Salmoniere Y;
WPI; 1999-181045/15.
P-PSDB; AA04750.
Mycobacterial DNA vectors containing reporter constructs - for
identifying coding or promoter sequences involved in infection-associated
protein expression.
Claim 22; Fig 1D; 309pp; French.
Sequences AAX34001-X34252 represent nucleic acids encoding secreted
proteins from various Mycobacterium species microorganisms. The
nucleotide sequences can be used as primers and probes for methods for
detecting and identifying mycobacteria, especially belonging to the M.
tuberculosis complex. The encoded proteins can be used in vaccines for
immunisation against a bacterial or viral infection

Sequence 411 BP; 68 A; 130 C; 146 G; 67 T; 0 U; 0 Other;

Query Match 100.0%; Score 30; DB 2; Length 411;
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAACGGCGAGGGGATCAGTACTGCGCAACA 30
 |||||
 DB 64 GCAACGGCGAGGGGATCAGTACTGCGCAACA 35

RESULT 2
 AAC97021/c
 ID AAC97021 standard; DNA; 411 BP.
 AC AAC97021;
 XX
 DT 23-FEB-2001 (first entry)
 XX
 DE Mycobacterium tuberculosis secreted protein cDNA #43.
 KW Mycobacterium tuberculosis secreted protein; MTSP; vaccine; ss.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO200066143-A1.
 XX
 PD 09-NOV-2000.
 XX
 PP 04-MAY-2000; 2000WO-US012197.
 XX
 PR 04-MAY-1999; 99US-0132479P.
 PR 04-MAY-1999; 99US-0132503P.
 XX
 PA (PUBL-) PUBLIC HEALTH RES INST NEW YORK.
 XX
 PI Gennaro ML, Gomez MJ;
 DR WPI; 2001-007151/01.
 XX
 PT Novel Mycobacterium tuberculosis secreted polypeptides and
 PT polynucleotides useful in diagnosis, treatment and prophylaxis of
 PT tuberculosis.
 XX
 PS Claim 1; Fig 2; 60pp; English.
 CC
 CC The present invention relates to Mycobacterium tuberculosis secreted
 CC proteins (MTSP), where the polypeptide has M. tuberculosis specific
 CC antigenic and immunogenic properties. Compositions of the invention may
 CC be useful for diagnosing Mycobacterium tuberculosis infection and as a
 CC vaccine against M. tuberculosis infection
 XX
 SQ Sequence 411 BP; 68 A; 130 C; 146 G; 67 T; 0 U; 0 Other;

Query Match 100.0%; Score 30; DB 4; Length 411;
 Best Local Similarity 100.0%; Pred. No. 0.0015;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAACGGCGAGGGGATCAGTACTGCGCAACA 30
 |||||
 DB 64 GCAACGGCGAGGGGATCAGTACTGCGCAACA 35

RESULT 3
 AAX34006/c
 ID AAX34006 standard; DNA; 498 BP.
 AC AAX34006;
 XX
 DT 06-JUL-1999 (first entry)
 XX
 DE Mycobacterium species nucleic acid sequence 1P.
 XX
 KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;

Query Match 100.0%; Score 30; DB 2; Length 498;
 Best Local Similarity 100.0%; Pred. No. 0.0016;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAACGGCGAGGGGATCAGTACTGCGCAACA 30
 |||||
 DB 151 GCAACGGCGAGGGGATCAGTACTGCGCAACA 122

RESULT 4
 AAX34004/c
 ID AAX34004 standard; DNA; 959 BP.
 AC AAX34004;
 XX
 DT 06-JUL-1999 (first entry)
 XX
 DE Mycobacterium species nucleic acid sequence 1C'.
 XX
 KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection; ss.
 XX
 OS Mycobacterium sp.
 XX
 PN WO9909186-A2.
 XX
 PD 25-FEB-1999.
 XX
 PF 14-AUG-1998; 98WO-FR001813.
 XX
 PR 14-AUG-1997; 97FR-00010404.
 PR 11-SEP-1997; 97FR-00011325.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Gicquel B, Portnoie D, Lim E, Pelicic V, Guigneno A;
 PI Goguet De La Salmoniere Y;

hybridisation; detection; vaccine; immunisation; infection; ss.
 XX
 OS Mycobacterium sp.
 XX
 PN WO9909186-A2.
 XX
 PD 25-FEB-1999.
 XX
 PF 14-AUG-1998; 98WO-FR001813.
 XX
 PR 14-AUG-1997; 97FR-00010404.
 PR 11-SEP-1997; 97FR-00011325.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Gicquel B, Portnoie D, Lim E, Pelicic V, Guigneno A;
 PI Goguet De La Salmoniere Y;
 XX
 DR WPI; 1999-181045/15.
 DR P-PSDB; AAY04751.
 XX
 PT Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 XX
 PS Claim 22; Fig 1F; 309pp; French.
 XX
 CC Sequences AAX34001-X34252 represent nucleic acids encoding secreted
 CC proteins from various Mycobacterium species microorganisms. The
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection
 XX
 SQ Sequence 498 BP; 87 A; 161 C; 173 G; 77 T; 0 U; 0 Other;

Query Match 100.0%; Score 30; DB 2; Length 498;
 Best Local Similarity 100.0%; Pred. No. 0.0016;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAACGGCGAGGGGATCAGTACTGCGCAACA 30
 |||||
 DB 151 GCAACGGCGAGGGGATCAGTACTGCGCAACA 122

XX WPI; 1999-181045/15.
DR P-PSDB; AAY04749.
XX
XX
PT Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in infection-associated
PT protein expression.
XX
XX Claim 22; Fig 1C'; 309pp; French.
XX
CC Sequences AAX34001-X34252 represent nucleic acids encoding secreted
CC proteins from various Mycobacterium species microorganisms. The
CC nucleotide sequences can be used as primers and probes for methods for
CC detecting and identifying mycobacteria, especially belonging to the M.
CC tuberculosis complex. The encoded proteins can be used in vaccines for
CC immunisation against a bacterial or viral infection
XX
XX Sequence 959 BP; 164 A; 313 C; 324 G; 158 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 30; DB 2; Length 959;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAACGCGAGGCGGATCAGTACTGCCAACA 30
DB 722 GCAACGCGAGGCGGATCAGTACTGCCAACA 693
RESULT 5
AAX34003/c
ID AAX34003 standard; DNA; 960 BP.
AC AAX34003;
XX
XX 06-JUL-1999 (first entry)
XX
XX Mycobacterium species nucleic acid sequence 1B'.
DE
KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection; ss.
XX
XX Mycobacterium sp.
OS
XX WO9909186-A2.
PN
XX 25-FEB-1999.
PD
XX 14-AUG-1998; 98WO-FR001813.
PF
XX 14-AUG-1997; 97FR-00010404.
PR
XX 11-SEP-1997; 97FR-00011325.
PR
XX (INSP) INST PASTEUR.
PA
XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
PI Goguet De La Salmoniere Y;
PI
XX WPI; 1999-181045/15.
DR
XX P-PSDB; AAY04748.
DR
XX WPI; 1999-181045/15.
DR
XX P-PSDB; AAY04749.
DR
XX Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in infection-associated
PT protein expression.
XX
XX Claim 22; Fig 1B'; 309pp; French.
XX
CC Sequences AAX34001-X34252 represent nucleic acids encoding secreted
CC proteins from various Mycobacterium species microorganisms. The
CC nucleotide sequences can be used as primers and probes for methods for
CC detecting and identifying mycobacteria, especially belonging to the M.
CC tuberculosis complex. The encoded proteins can be used in vaccines for
CC immunisation against a bacterial or viral infection
XX

SQ Sequence 960 BP; 165 A; 313 C; 324 G; 158 T; 0 U; 0 Other;
Query Match 100.0%; Score 30; DB 2; Length 960;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAACGCGAGGCGGATCAGTACTGCCAACA 30
DB 723 GCAACGCGAGGCGGATCAGTACTGCCAACA 694
RESULT 6
AAX34002/c
ID AAX34002 standard; DNA; 961 BP.
XX
XX AC AAX34002;
XX
XX 06-JUL-1999 (first entry)
XX
XX Mycobacterium species nucleic acid sequence 1A'.
DE
XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KW hybridisation; detection; vaccine; immunisation; infection; ss.
XX
XX Mycobacterium sp.
OS
XX WO9909186-A2.
PN
XX 25-FEB-1999.
PD
XX 14-AUG-1998; 98WO-FR001813.
PF
XX 14-AUG-1997; 97FR-00010404.
PR
XX 11-SEP-1997; 97FR-00011325.
PR
XX (INSP) INST PASTEUR.
PA
XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
PI Goguet De La Salmoniere Y;
PI
XX WPI; 1999-181045/15.
DR
XX P-PSDB; AAY04747.
DR
XX Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in infection-associated
PT protein expression.
XX
XX Claim 22; Fig 1A'; 309pp; French.
XX
CC Sequences AAX34001-X34252 represent nucleic acids encoding secreted
CC proteins from various Mycobacterium species microorganisms. The
CC nucleotide sequences can be used as primers and probes for methods for
CC detecting and identifying mycobacteria, especially belonging to the M.
CC tuberculosis complex. The encoded proteins can be used in vaccines for
CC immunisation against a bacterial or viral infection
XX
XX Sequence 961 BP; 165 A; 313 C; 325 G; 158 T; 0 U; 0 Other;
SQ
Query Match 100.0%; Score 30; DB 2; Length 961;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCAACGCGAGGCGGATCAGTACTGCCAACA 30
DB 724 GCAACGCGAGGCGGATCAGTACTGCCAACA 695
RESULT 7
AAX34007/c
ID AAX34007 standard; DNA; 1063 BP.
XX
XX AC AAX34007;
XX

```

DT XX 06-JUL-1999 (first entry)
DE XX Mycobacterium species nucleic acid sequence 2.
XX
XX KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
XX KW Hybridisation; detection; vaccine; immunisation; infection; ss.
XX
XX OS Mycobacterium sp.
XX PN WO9909186-A2.
XX
XX PD 25-FEB-1999.
XX
XX PF 14-AUG-1998; 98WO-FR001813.
XX
XX PR 14-AUG-1997; 97FR-00010404.
XX PR 11-SEP-1997; 97FR-00011325.
XX
XX PA (INSP ) INST PASTEUR.
XX
XX PI Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
XX PI Goguet De La Salmoniere Y;
XX
XX DR WPI; 1999-181045/15.
XX DR P-PSDB; AAY04752.
XX
XX PT Mycobacterial DNA vectors containing reporter constructs - for
XX PT identifying coding or promoter sequences involved in infection-associated
XX PT protein expression.
XX
XX PS Claim 22; Fig 2; 309pp; French.
XX
XX CC Sequences AAX34001-X34252 represent nucleic acids encoding secreted
XX CC proteins from various Mycobacterium species microorganisms. The
XX CC nucleotide sequences can be used as primers and probes for methods for
XX CC detecting and identifying mycobacteria, especially belonging to the M.
XX CC tuberculosis complex. The encoded proteins can be used in vaccines for
XX CC immunisation against a bacterial or viral infection
XX
XX SQ Sequence 1063 BP; 173 A; 343 C; 374 G; 173 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 30; DB 2; Length 1063;
XX Best Local Similarity 100.0%; Pred. No. 0.0017;
XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GCAACGCGAGGCGATCAGTACTGCCAACA 30
XX |||||
XX Db 514 GCAACGCGAGGCGATCAGTACTGCCAACA 485
XX
XX RESULT 8
XX AAX34001/c
XX ID AAX34001 standard; DNA; 1243 BP.
XX
XX AC AAX34001;
XX
XX XX
XX DT 06-JUL-1999 (first entry)
XX
XX DE Mycobacterium species nucleic acid sequence 1.
XX
XX KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
XX KW Hybridisation; detection; vaccine; immunisation; infection; ss.
XX
XX OS Mycobacterium sp.
XX PN WO9909186-A2.
XX
XX PD 25-FEB-1999.
XX
XX PF 14-AUG-1998; 98WO-FR001813.
XX
XX PR 14-AUG-1997; 97FR-00010404.
XX PR 11-SEP-1997; 97FR-00011325.
XX
XX PA (INSP ) INST PASTEUR.
XX
XX PI Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
XX PI Goguet De La Salmoniere Y;
XX
XX DR WPI; 1999-181045/15.
XX DR P-PSDB; AAY04752.
XX
XX PT Mycobacterial DNA vectors containing reporter constructs - for
XX PT identifying coding or promoter sequences involved in infection-associated
XX PT protein expression.
XX
XX PS Claim 22; Fig 2; 309pp; French.
XX
XX CC Sequences AAX34001-X34252 represent nucleic acids encoding secreted
XX CC proteins from various Mycobacterium species microorganisms. The
XX CC nucleotide sequences can be used as primers and probes for methods for
XX CC detecting and identifying mycobacteria, especially belonging to the M.
XX CC tuberculosis complex. The encoded proteins can be used in vaccines for
XX CC immunisation against a bacterial or viral infection
XX
XX SQ Sequence 1063 BP; 173 A; 343 C; 374 G; 173 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 30; DB 2; Length 1063;
XX Best Local Similarity 100.0%; Pred. No. 0.0017;
XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GCAACGCGAGGCGATCAGTACTGCCAACA 30
XX |||||
XX Db 514 GCAACGCGAGGCGATCAGTACTGCCAACA 485
XX
XX RESULT 8
XX AAX34001/c
XX ID AAX34001 standard; DNA; 1243 BP.
XX
XX AC AAX34001;
XX
XX XX
XX DT 06-JUL-1999 (first entry)
XX
XX DE Mycobacterium species nucleic acid sequence 1.
XX
XX KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
XX KW Hybridisation; detection; vaccine; immunisation; infection; ss.
XX
XX OS Mycobacterium sp.
XX PN WO9909186-A2.
XX
XX PD 25-FEB-1999.
XX
XX PF 14-AUG-1998; 98WO-FR001813.
XX
XX PR 14-AUG-1997; 97FR-00010404.
XX PR 11-SEP-1997; 97FR-00011325.
XX

```

```

XX (INSP ) INST PASTEUR.
XX
XX Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
XX Goguet De La Salmoniere Y;
XX
XX WPI; 1999-181045/15.
XX P-PSDB; AAY04742.
XX
XX Mycobacterial DNA vectors containing reporter constructs - for
XX PT identifying coding or promoter sequences involved in infection-associated
XX PT protein expression.
XX
XX PS Claim 22; Fig 1; 309pp; French.
XX
XX CC Sequences AAX34001-X34252 represent nucleic acids encoding secreted
XX CC proteins from various Mycobacterium species microorganisms. The
XX CC nucleotide sequences can be used as primers and probes for methods for
XX CC detecting and identifying mycobacteria, especially belonging to the M.
XX CC tuberculosis complex. The encoded proteins can be used in vaccines for
XX CC immunisation against a bacterial or viral infection
XX
XX SQ Sequence 1243 BP; 218 A; 394 C; 422 G; 209 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 30; DB 2; Length 1243;
XX Best Local Similarity 100.0%; Pred. No. 0.0018;
XX Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GCAACGCGAGGCGATCAGTACTGCCAACA 30
XX |||||
XX Db 1005 GCAACGCGAGGCGATCAGTACTGCCAACA 976
XX
XX RESULT 9
XX AAI99682_02/c
XX Continuation (3 of 45) of AAI99682 from base 200001 (Mycobacterium tuberculosis strain H:
XX WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682
XX
XX Fragment Name Begin End
XX WP AAI99682_00 1 110000
XX WP AAI99682_01 100001 210000
XX WP AAI99682_02 200001 310000
XX WP AAI99682_03 300001 410000
XX WP AAI99682_04 400001 510000
XX WP AAI99682_05 500001 610000
XX WP AAI99682_06 600001 710000
XX WP AAI99682_07 700001 810000
XX WP AAI99682_08 800001 910000
XX WP AAI99682_09 900001 1010000
XX WP AAI99682_10 1000001 1110000
XX WP AAI99682_11 1100001 1210000
XX WP AAI99682_12 1200001 1310000
XX WP AAI99682_13 1300001 1410000
XX WP AAI99682_14 1400001 1510000
XX WP AAI99682_15 1500001 1610000
XX WP AAI99682_16 1600001 1710000
XX WP AAI99682_17 1700001 1810000
XX WP AAI99682_18 1800001 1910000
XX WP AAI99682_19 1900001 2010000
XX WP AAI99682_20 2000001 2110000
XX WP AAI99682_21 2100001 2210000
XX WP AAI99682_22 2200001 2310000
XX WP AAI99682_23 2300001 2410000
XX WP AAI99682_24 2400001 2510000
XX WP AAI99682_25 2500001 2610000
XX WP AAI99682_26 2600001 2710000
XX WP AAI99682_27 2700001 2810000
XX WP AAI99682_28 2800001 2910000
XX WP AAI99682_29 2900001 3010000
XX WP AAI99682_30 3000001 3110000
XX WP AAI99682_31 3100001 3210000
XX WP AAI99682_32 3200001 3310000
XX WP AAI99682_33 3300001 3410000
XX WP AAI99682_34 3400001 3510000

```

```
WP AA199682_35 350001 3610000
WP AA199682_36 360001 3710000
WP AA199682_37 370001 3810000
WP AA199682_38 380001 3910000
WP AA199682_39 390001 4010000
WP AA199682_40 400001 4110000
WP AA199682_41 410001 4210000
WP AA199682_42 420001 4310000
WP AA199682_43 430001 4410000
WP AA199682_44 440001 4411529

Query Match 100.0%; Score 30; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.0032;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAACGCGAGGCGGATCAGTACTGCGCAACA 30
Db 41575 GCAACGCGAGGCGGATCAGTACTGCGCAACA 41546

RESULT 10
AA199683_02/c
Continuation (3 of 44) of AA199683 from base 200001 (Mycobacterium tuberculosis strain H
WP Sequence split into 44 fragments LOCUS AA199683 Accession AA199683
WP Fragment Name Begin End
WP AA199683_00 1 110000
WP AA199683_01 100001 210000
WP AA199683_02 200001 310000
WP AA199683_03 300001 410000
WP AA199683_04 400001 510000
WP AA199683_05 500001 610000
WP AA199683_06 600001 710000
WP AA199683_07 700001 810000
WP AA199683_08 800001 910000
WP AA199683_09 900001 1010000
WP AA199683_10 1000001 1110000
WP AA199683_11 1100001 1210000
WP AA199683_12 1200001 1310000
WP AA199683_13 1300001 1410000
WP AA199683_14 1400001 1510000
WP AA199683_15 1500001 1610000
WP AA199683_16 1600001 1710000
WP AA199683_17 1700001 1810000
WP AA199683_18 1800001 1910000
WP AA199683_19 1900001 2010000
WP AA199683_20 2000001 2110000
WP AA199683_21 2100001 2210000
WP AA199683_22 2200001 2310000
WP AA199683_23 2300001 2410000
WP AA199683_24 2400001 2510000
WP AA199683_25 2500001 2610000
WP AA199683_26 2600001 2710000
WP AA199683_27 2700001 2810000
WP AA199683_28 2800001 2910000
WP AA199683_29 2900001 3010000
WP AA199683_30 3000001 3110000
WP AA199683_31 3100001 3210000
WP AA199683_32 3200001 3310000
WP AA199683_33 3300001 3410000
WP AA199683_34 3400001 3510000
WP AA199683_35 3500001 3610000
WP AA199683_36 3600001 3710000
WP AA199683_37 3700001 3810000
WP AA199683_38 3800001 3910000
WP AA199683_39 3900001 4010000
WP AA199683_40 4000001 4110000
WP AA199683_41 4100001 4210000
WP AA199683_42 4200001 4310000
WP AA199683_43 4300001 4403765

Query Match 100.0%; Score 30; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.0032;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 GCAACGCGAGGCGGATCAGTACTGCGCAACA 30
Db 41686 GCAACGCGAGGCGGATCAGTACTGCGCAACA 41657

RESULT 11
AAZ01425_03
Continuation (4 of 11) of AAZ01425 from base 300001 (Complete genome sequence of Chlamydi
WP Sequence split into 11 fragments LOCUS AAZ01425 Accession AAZ01425
WP Fragment Name Begin End
WP AAZ01425_00 1 110000
WP AAZ01425_01 100001 210000
WP AAZ01425_02 200001 310000
WP AAZ01425_03 300001 410000
WP AAZ01425_04 400001 510000
WP AAZ01425_05 500001 610000
WP AAZ01425_06 600001 710000
WP AAZ01425_07 700001 810000
WP AAZ01425_08 800001 910000
WP AAZ01425_09 900001 1010000
WP AAZ01425_10 1000001 1038602

Query Match 64.7%; Score 19.4; DB 2; Length 110000;
Best Local Similarity 79.3%; Pred. No. 1.7e+02;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CAAACGCGAGGCGGATCAGTACTGCGCAACA 30
Db 25973 CAAAGTTAGGCGGATCAGTACTGCGCAACA 26001

RESULT 12
AAD03951
ID AAD03951 standard; DNA; 3328 BP.
XX AC AAD03951;
XX DT 11-SEP-2003 (revised)
XX DT 02-JUL-2001 (first entry)
XX DE A. nidulans alpha-1,2 mannosidase 1A gene.
XX KW Alpha-1,2 mannosidase; glycosylation pattern; glycoprotein; ds.
XX OS Emericella nidulans.
FH Key Location/Qualifiers
FT misc_signal 536..540
FT TATA_signal 660..665
FT CDS 706..3203
FT /tag= a
FT /note= "CCATT upstream element"
FT /tag= b
FT /tag= c
FT /product= "A. nidulans alpha-1,2 mannosidase 1A protein"
FT /transl_except= (pos:1603..1608, aa: Ala Lys Arg Ser Ala
FT Gly Pro Glu Lys Gly Asn Trp His Val Val Ala Thr Asp)
FT /transl_except= (pos:2623..2676, aa:Asn Ala)
FT /transl_except= (pos:3100..3102, aa:Asp)
FT /transl_except= (pos:3195..3197, aa:Lys)
FT /note= "CDS is interrupted by 1 intron"
FT /tag= d
FT /tag= e
FT /number= 1
FT /tag= f
FT /number= 1
FT /tag= g
FT /note= "5' splice site of intron 1"
FT /tag= g
```

FT misc_feature /note= "Internal lariat sequence of intron 1"
 FT 3161. .3163
 FT /*tag= h
 FT /note= "3' splice site of intron"
 FT 3164. .3203
 FT /*tag= i
 FT /number= 2
 FT exon
 FT WO200125406-A1.
 FT 12-APR-2001.
 FT 02-OCT-2000; 2000WO-US027210.
 FT 01-OCT-1999; 99US-0157341P.
 FT (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 FT Hintz WE, Bades CJ;
 FT WPI; 2001-290610/30.
 FT P-PSDB; AAE00646.
 FT New mannosidase enzymes, useful for modifying or altering the
 FT glycosylation patterns of macromolecules, particularly for modifying
 FT target protein.
 FT Claim 11a; Fig 1; 87pp; English.

XX The invention relates to Aspergillus nidulans alpha-1,2 mannosidase genes
 CC and proteins. The mannosidase enzymes are useful for modifying or
 CC altering the glycosylation patterns of macromolecules, e.g. proteins. In
 CC particular, the enzymes useful for modifying target proteins. This
 CC enables glycoproteins to be engineered to be more effectively used and
 CC produced. The present sequence is A. nidulans mannosidase 1A gene. Note:
 CC This sequence is stated as being the same as that shown as SEQ ID NO:1
 CC (AAD03592) in the sequence listing of the specification. However the
 CC sequences differ at several positions. (Updated on 11-SEP-2003 to
 CC standardise OS field)

XX Sequence 3328 BP; 809 A; 925 C; 811 G; 783 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 4; Length 3328;
 Best Local Similarity 81.5%; Pred. No. 1.6e+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AACGCGAGGGGATCAGTACTGCGAAC 29
 DB 86 AACGCGGGGCAATCAGTTCAACCAAC 112

RESULT 13

AAD03952
 ID AAD03952 standard; DNA; 3328 BP.

XX AAD03952;

XX 11-SEP-2003 (revised)

DT 02-JUL-2001 (first entry)

DE Alternative version of A. nidulans alpha-1,2 mannosidase 1A gene.

XX Alpha-1,2 mannosidase; glycosylation pattern; glycoprotein; ds.

XX Emericella nidulans.

XX Key Location/Qualifiers

FT misc_signal 536..540

FT /*tag= a

FT TATA_signal /note= "CCATT upstream element"

FT 660..665

FT /*tag= b

FT 706..3203

FT CDS

FT exon
 FT 706..3113
 FT /*tag= d
 FT /number= 1
 FT 3114..3163
 FT /*tag= e
 FT /number= 1
 FT 3114..3119
 FT /tag= f
 FT /note= "5' splice site of intron 1"
 FT 3147..3151
 FT /*tag= g
 FT /note= "Internal lariat sequence of intron 1"
 FT 3161..3163
 FT /*tag= h
 FT /note= "3' splice site of intron 1"
 FT 3164..3203
 FT /*tag= i
 FT /number= 2
 FT exon
 FT WO200125406-A1.
 FT 12-APR-2001.
 FT 02-OCT-2000; 2000WO-US027210.
 FT 01-OCT-1999; 99US-0157341P.
 FT (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 FT Hintz WE, Bades CJ;
 FT WPI; 2001-290610/30.
 FT P-PSDB; AAE00647.
 FT New mannosidase enzymes, useful for modifying or altering the
 FT glycosylation patterns of macromolecules, particularly for modifying
 FT target protein.
 FT Claim 11a; Page 71-72; 87pp; English.

XX The invention relates to Aspergillus nidulans alpha-1,2 mannosidase genes
 CC and proteins. The mannosidase enzymes are useful for modifying or
 CC altering the glycosylation patterns of macromolecules, e.g. proteins. In
 CC particular, the enzymes useful for modifying target proteins. This
 CC enables glycoproteins to be engineered to be more effectively used and
 CC produced. The present sequence is an alternative version of A. nidulans
 CC mannosidase 1A gene. Note: This sequence is stated as being the same as
 CC that shown in Fig 1 (AAD03591) of the specification. However the
 CC sequences differ at several positions. (Updated on 11-SEP-2003 to
 CC standardise OS field)

XX Sequence 3328 BP; 809 A; 927 C; 809 G; 783 T; 0 U; 0 Other;

Query Match 63.3%; Score 19; DB 4; Length 3328;

Best Local Similarity 81.5%; Pred. No. 1.6e+02;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 AACGCGAGGGGATCAGTACTGCGAAC 29

DB 86 AACGCGGGGCAATCAGTTCAACCAAC 112

RESULT 14

ABL28961

ID ABL28961 standard; DNA; 1163 BP.

XX ABL28961;

AC ABL28961;

```
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 38356.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PP 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX FR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX PS Claim 1; SEQ ID NO 38356; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1163 BP; 270 A; 302 C; 341 G; 250 T; 0 U; 0 Other;

Query Match 62.7%; Score 18.8; DB 4; Length 1163;
Best Local Similarity 90.9%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCAACGCGAGCGGATCAGTAC 22
DB 800 GAAACGCGAGCGGATCAGTAC 821

RESULT 15
ABL28969
ID ABL28969 standard; DNA; 1166 BP.
XX AC ABL28969;
XX AC ABL28969;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 38380.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ds.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PP 23-MAR-2001; 2001WO-US009231.
```

```
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX PS Claim 1; SEQ ID NO 38380; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1166 BP; 273 A; 298 C; 335 G; 260 T; 0 U; 0 Other;

Query Match 62.7%; Score 18.8; DB 4; Length 1166;
Best Local Similarity 90.9%; Pred. No. 1.7e+02;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GCAACGCGAGCGGATCAGTAC 22
DB 803 GAAACGCGAGCGGATCAGTAC 824

Search completed: April 29, 2004, 04:56:57
Job time : 250.747 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:13:15 ; Search time 52.9412 Seconds
(without alignments)
314.472 Million cell updates/sec

Title: US-10-624-714-4
Perfect score: 30
Sequence: 1 gcaacgcgagggcgatcagctactgccaaca 30

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A COMB.seq*
2: /cgn2_6/prodata/2/ina/5B COMB.seq*
3: /cgn2_6/prodata/2/ina/6A COMB.seq*
4: /cgn2_6/prodata/2/ina/6B COMB.seq*
5: /cgn2_6/prodata/2/ina/PCUTUS COMB.seq*
6: /cgn2_6/prodata/2/ina/backfiles1.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	30	100.0	4403765	3	US-09-103-840A-2
C 2	30	100.0	4411529	3	US-09-103-840A-1
C 3	18.8	62.7	777	4	US-09-252-991A-12518
C 4	18.8	62.7	846	4	US-09-252-991A-12955
C 5	18.8	62.7	1755	4	US-09-252-991A-13978
C 6	18.8	62.7	2280	4	US-09-252-991A-12824
C 7	18.6	62.0	684	3	US-08-911-853-22
C 8	18.6	62.0	684	3	US-09-479-409-22
C 9	18.6	62.0	684	4	US-09-479-453-22
C 10	18.6	62.0	17612	3	US-08-911-853-29
C 11	18.6	62.0	17612	3	US-08-479-409-29
C 12	18.6	62.0	17612	4	US-09-479-453-29
C 13	18.4	61.3	261	4	US-09-489-039A-1139
C 14	18.4	61.3	429	4	US-09-489-039A-1211
C 15	18.4	61.3	564	4	US-09-489-039A-1112
C 16	18.4	61.3	741	4	US-09-489-039A-994
C 17	18	60.0	1314	4	US-08-252-991A-13338
C 18	18	60.0	1359	4	US-09-252-991A-12612
C 19	17.8	59.3	1362	4	US-09-489-039A-1745
C 20	17.8	59.3	1377	4	US-09-489-039A-6783
C 21	17.8	59.3	1425	4	US-09-489-039A-3261
C 22	17.8	59.3	68750	3	US-09-335-409-1
C 23	17.8	59.3	68750	4	US-09-568-102-1
C 24	17.8	59.3	68750	4	US-08-567-969-1
C 25	17.8	59.3	68750	4	US-09-568-480-1
C 26	17.8	59.3	68750	4	US-09-568-486-1
C 27	17.8	59.3	68750	4	US-09-568-472-1

C 28	17.8	59.3	68750	4	US-09-567-899-1	Sequence 1, Appli
C 29	17.8	59.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 30	17.8	59.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 31	17.6	58.7	14561	4	US-09-392-714-1	Sequence 1, Appli
C 32	17.4	58.0	474	4	US-09-252-991A-6386	Sequence 371, App
C 33	17.4	58.0	741	4	US-09-401-064-371	Sequence 202, App
C 34	17.4	58.0	810	4	US-09-252-991A-214	Sequence 214, App
C 35	17.4	58.0	1077	4	US-09-252-991A-6161	Sequence 6161, App
C 36	17.4	58.0	1173	4	US-09-252-991A-6004	Sequence 6004, App
C 37	17.4	58.0	1494	4	US-09-252-991A-206	Sequence 1, Appli
C 38	17.4	58.0	1890	4	US-09-214-808-1	Sequence 5692, App
C 39	17.4	58.0	536165	4	US-09-489-039A-5692	Sequence 12111, A
C 40	17.2	57.3	294	4	US-09-252-991A-12111	Sequence 12250, A
C 41	17.2	57.3	525	4	US-09-252-991A-12111	Sequence 1510, App
C 42	17.2	57.3	924	4	US-09-252-991A-1510	Sequence 5866, App
C 43	17.2	57.3	927	4	US-09-252-991A-1510	Sequence 12311, A
C 44	17.2	57.3	945	4	US-09-489-039A-5866	
C 45	17.2	57.3	1011	4	US-09-252-991A-12311	

ALIGNMENTS

RESULT 1
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 100.0%; Score 30; DB 3; Length 4403765;
Best Local Similarity 100.0%; Pred. No. 0.0006;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAACGCGAGGGCGGATCAGTACTGCGCAACA 30
DB 241686 GCAACGCGAGGGCGGATCAGTACTGCGCAACA 241657

RESULT 2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103.840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 4411529

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; OTHER INFORMATION: H37RV

US-09-103-840A-1

Query Match 100.0%; Score 30; DB 3; Length 4411529;

Best Local Similarity 100.0%; Pred. No. 0.0006;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCAACGGAGGGCGATCAGTACTGCCAACA 30

DB 241575 GCAACGGAGGGCGATCAGTACTGCCAACA 241546

RESULT 3

US-09-252-991A-12518/c

; Sequence 12518, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 12518

; LENGTH: 777

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12518

Query Match 62.7%; Score 18.8; DB 4; Length 777;

Best Local Similarity 76.7%; Pred. No. 16;

Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCAACGGAGGGCGATCAGTACTGCCAACA 30

DB 172 GCATCGGAGCGCGATCAGCAGGCGGAACA 143

RESULT 4

US-09-252-991A-12955

; Sequence 12955, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 12955

; LENGTH: 846

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12955

Query Match 62.7%; Score 18.8; DB 4; Length 846;

Best Local Similarity 76.7%; Pred. No. 16;

Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCAACGGAGGGCGATCAGTACTGCCAACA 30

DB 636 GCATCGGAGCGCGATCAGCAGGCGGAACA 665

RESULT 5

US-09-543-681A-3978/c

; Sequence 3978, Application US/09543681A

; Patent No. 6605709

; GENERAL INFORMATION:

; APPLICANT: GARY BRETTON

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.1002-001

; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05

; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 3978

; LENGTH: 1755

; TYPE: DNA

; ORGANISM: Proteus mirabilis

US-09-543-681A-3978

Query Match 62.7%; Score 18.8; DB 4; Length 1755;

Best Local Similarity 76.7%; Pred. No. 19;

Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCAACGGAGGGCGATCAGTACTGCCAACA 30

DB 1199 GCAACGAGAGCGCGCTCAATATTGCCAACA 1170

RESULT 6

US-09-252-991A-12824/c

; Sequence 12824, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 12824

; LENGTH: 2280

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-12824

Query Match 62.7%; Score 18.8; DB 4; Length 2280;

Best Local Similarity 76.7%; Pred. No. 20;

Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 GCAACGGAGGGCGATCAGTACTGCCAACA 30

DB 1622 GCATCGGAGCGCGATCAGCAGGCGGAACA 1593

RESULT 7

US-08-911-853-22/c

; Sequence 22, Application US/08911853

; Patent No. 6048710

; GENERAL INFORMATION:

; APPLICANT: Gerritse, Gijbert

; APPLICANT: Quax, Wilhelmus J.

```

; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 684 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-479-409-22

Query Match 62.0%; Score 18.6; DB 3; Length 684;
Best Local Similarity 84.0%; Pred. No.19;
Matches 21; Conservative 0; Mismatches 4; Indels

QY 6 GCGAGGCGCATGACTGCCAACA 30
DB 47 GCGATGCGCATGACGACTCCAGCA 23

RESULT 9
US-09-479-453-22/c
; Sequence 22, Application US/09479453
; Patent No. 6313283
; GENERAL INFORMATION:
; APPLICANT: Gerritse, Gilsbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/479,453
; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/911,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 684 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-479-453-22

Query Match 62.0%; Score 18.6; DB 4; Length 684;
Best Local Similarity 84.0%; Pred. No.19;
Matches 21; Conservative 0; Mismatches 4; Indels

QY 6 GCGAGGCGCATGACTGCCAACA 30

```

Db 47 GCGATGGCGATCAGCACTTCCAGCA 23

RESULT 10

US-08-911-853-29/c

Sequence 29, Application US/08911853

Patent No. 6048710

GENERAL INFORMATION:

APPLICANT: Gerritse, Gijbert

APPLICANT: Quax, Wilhelmus J.

TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED

TITLE OF INVENTION: EXPRESSION LEVELS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genencor International

STREET: 925 Page Mill Road

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1013

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/911,853

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/699,092

FILING DATE: 16-AUG-1996

ATTORNEY/AGENT INFORMATION:

NAME: Glaister, Debra J

REGISTRATION NUMBER: 33,888

REFERENCE/DOCKET NUMBER: GC361-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-846-7620

TELEFAX: 650-845-6504

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 17612 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-911-853-29

Query Match 62.0%; Score 18.6; DB 3; Length 17612;

Best Local Similarity 84.0%; Pred. No. 35;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 GCGAGGGCGATCAGTACTGCCAACA 30

Db 14235 GCGATGGCGATCAGCACTTCCAGCA 14211

RESULT 11

US-09-479-409-29/c

Sequence 29, Application US/09479409

Patent No. 6225106

GENERAL INFORMATION:

APPLICANT: Gerritse, Gijbert

APPLICANT: Quax, Wilhelmus J.

TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED

TITLE OF INVENTION: EXPRESSION LEVELS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genencor International

STREET: 925 Page Mill Road

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1013

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/479,453

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/911,853

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Glaister, Debra J

REGISTRATION NUMBER: 33,888

REFERENCE/DOCKET NUMBER: GC361-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-846-7620

TELEFAX: 650-845-6504

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 17612 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-911-853-29

Query Match 62.0%; Score 18.6; DB 3; Length 17612;

Best Local Similarity 84.0%; Pred. No. 35;

Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 6 GCGAGGGCGATCAGTACTGCCAACA 30

Db 14235 GCGATGGCGATCAGCACTTCCAGCA 14211

RESULT 12

US-09-479-453-29/c

Sequence 29, Application US/09479453

Patent No. 6313283

GENERAL INFORMATION:

APPLICANT: Gerritse, Gijbert

APPLICANT: Quax, Wilhelmus J.

TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED

TITLE OF INVENTION: EXPRESSION LEVELS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genencor International

STREET: 925 Page Mill Road

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1013

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/479,453

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/911,853

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Glaister, Debra J

REGISTRATION NUMBER: 33,888

REFERENCE/DOCKET NUMBER: GC361-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-846-7620

TELEFAX: 650-845-6504

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 17612 base pairs

TYPE: nucleic acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-479-453-29

  Query Match      62.0%; Score 18.6; DB 4; Length 17612;
Best Local Similarity 84.0%; Pred. No. 35;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      6 GCGAGGGCGGATCAGTACTGCGCAACA 30
Db      14235 GCGATGGCGGATCAGACTTCCAGCA 14211

RESULT 13
US-09-489-039A-1139
; Sequence 1139, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1139
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1139

  Query Match      61.3%; Score 18.4; DB 4; Length 261;
Best Local Similarity 78.6%; Pred. No. 20;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      2 CAACGGAGGGCGGATCAGTACTGCGCAAC 29
Db      167 CAACGGAGGGCTATCAGCGCGGCAGC 194

RESULT 14
US-09-489-039A-1211/c
; Sequence 1211, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1211
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1211

  Query Match      61.3%; Score 18.4; DB 4; Length 429;
Best Local Similarity 78.6%; Pred. No. 22;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      2 CAACGGAGGGCGGATCAGTACTGCGCAAC 29
Db      352 CAACGGAGGGCTATCAGCGCGGCAGC 325

RESULT 15
US-09-489-039A-1112
; Sequence 1112, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 1112
; LENGTH: 564
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-1112

  Query Match      61.3%; Score 18.4; DB 4; Length 564;
Best Local Similarity 78.6%; Pred. No. 23;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      2 CAACGGAGGGCGGATCAGTACTGCGCAAC 29
Db      64 CAACGGAGGGCTATCAGCGCGGCAGC 91

Search completed: April 29, 2004, 11:45:29
Job time : 67.9412 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:23:09 ; Search time 231.137 Seconds
(without alignments)
514.627 Million cell updates/sec

Title: US-10-624-714-5
Perfect score: 28
Sequence: 1 gctgggcagctcacagctacgactc 28

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseqn_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001s:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28	100.0	110000	4	AAI99682_04
2	28	100.0	110000	4	AAI99683_04
C 3	18.4	65.7	5746	4	ABL01952 Drosophil
C 4	18.4	65.7	147708	6	ABQ88154
C 5	18	64.3	356	6	ABL79024
C 6	18	64.3	386	4	AAI93288 Human pol
C 7	18	64.3	424	4	AAI85519 Human pol
C 8	18	64.3	473	8	ACH21700
C 9	18	64.3	610	2	AZ232083
C 10	18	64.3	1014	6	ABQ30232 M. capsul
C 11	18	64.3	2215	5	ABAI8457 Human ner
C 12	18	64.3	2215	5	ABAI8458 Human ner
C 13	18	64.3	6455	6	ABK84151 Human cDN
C 14	18	64.3	6455	9	ACF79923 Breast ca
C 15	18	64.3	6915	9	ADE62448 Human gen
C 16	18	64.3	6915	9	ADE62444 Human gen
C 17	18	64.3	6915	9	ADD46210 Human gen
C 18	18	64.3	7080	5	AS855528
C 19	17.6	62.9	112414	6	ABL59091
C 20	17.4	62.1	692	8	ADA48171
C 21	17.4	62.1	1152	5	AS81793
C 22	17.4	62.1	1704	7	ACA45452
C 23	17.4	62.1	2726	6	ABL65806 Lung canc

C 24	17.4	62.1	2858	4	AAI17381	Aad17381 Human bon
C 25	17.4	62.1	3082	4	AAI17400	Aad17400 Human bon
C 26	17.4	62.1	3082	6	AAI94967	Aas94967 Human DNA
C 27	17.4	62.1	3242	4	AAI17360	Aad17360 Human bon
C 28	17.4	62.1	3649	5	AAI13630	Aas13630 cDNA sequ
C 29	17.4	62.1	3808	2	AAV04634	Aav04634 Porcine p
C 30	17.4	62.1	3808	2	AAV74100	Aav74100 Porcine G
C 31	17.4	62.1	3808	3	AAZ86813	Aaz86813 Pig p120
C 32	17.4	62.1	4134	2	AAI58545	Aat58545 Human pho
C 33	17.4	62.1	4137	2	AAI58546	Aat58546 Human pho
C 34	17.4	62.1	5162	2	AAV74104	Aav74104 Human G-p
C 35	17.4	62.1	5162	3	AAZ86814	Aaz86814 Human p12
C 36	17.4	62.1	5397	6	AAI14367	Aas14367 cDNA enco
C 37	17.4	62.1	5397	6	ABL59522	ABL59522 Human pho
C 38	17.4	62.1	134525	2	AAQ04525	AAQ04525 Total bas
C 39	17.2	61.4	60	6	ABN41498	ABn41498 Human spl
C 40	17.2	61.4	1148	2	AAI71230	Aat71230 Papaya ri
C 41	17.2	61.4	2064	7	ACA25360	Aca25360 Prokaryot
C 42	17.2	61.4	3094	3	AAI98992	Aac98992 Human pan
C 43	17.2	61.4	3187	5	AAI64760	Aas64760 DNA enco
C 44	17.2	61.4	3442	2	AAI79183	Aax79183 Human cho
C 45	17.2	61.4	4687	3	AAA08582	Aaa08582 Human cyt

ALIGNMENTS

RESULT 1

AAI99682_04
Continuation (5 of 45) of AAI99682 from base 400001 (Mycobacterium tuberculosis strain H: WP Sequence split into 45 fragments LOCUS AAI99682 Accession Aai99682

WP	Fragment Name	Begin	End
WP	AAI99682_00	1	110000
WP	AAI99682_01	100001	210000
WP	AAI99682_02	200001	310000
WP	AAI99682_03	300001	410000
WP	AAI99682_04	400001	510000
WP	AAI99682_05	500001	610000
WP	AAI99682_06	600001	710000
WP	AAI99682_07	700001	810000
WP	AAI99682_08	800001	910000
WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000

```

WP AAI99682_41 410001 4210000
WP AAI99682_42 420001 4310000
WP AAI99682_43 430001 4410000
WP AAI99682_44 440001 4411529

Query Match 100.0%; Score 28; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACACAGTACGACTC 28
Db 72658 GCTGTGGCGCAGCTACACAGTACGACTC 72655

RESULT 2
AAI99683_04
Continuation (5 of 44) of AAI99683 from base 400001 (Mycobacterium tuberculosis strain H
WP Sequence split into 44 fragments LOCUS AAI99683 Accession Aai99683
WP Fragment Name Begin End
WP AAI99683_00 1 110000
WP AAI99683_01 100001 210000
WP AAI99683_02 200001 310000
WP AAI99683_03 300001 410000
WP AAI99683_04 400001 510000
WP AAI99683_05 500001 610000
WP AAI99683_06 600001 710000
WP AAI99683_07 700001 810000
WP AAI99683_08 800001 910000
WP AAI99683_09 900001 1010000
WP AAI99683_10 1000001 1110000
WP AAI99683_11 1100001 1210000
WP AAI99683_12 1200001 1310000
WP AAI99683_13 1300001 1410000
WP AAI99683_14 1400001 1510000
WP AAI99683_15 1500001 1610000
WP AAI99683_16 1600001 1710000
WP AAI99683_17 1700001 1810000
WP AAI99683_18 1800001 1910000
WP AAI99683_19 1900001 2010000
WP AAI99683_20 2000001 2110000
WP AAI99683_21 2100001 2210000
WP AAI99683_22 2200001 2310000
WP AAI99683_23 2300001 2410000
WP AAI99683_24 2400001 2510000
WP AAI99683_25 2500001 2610000
WP AAI99683_26 2600001 2710000
WP AAI99683_27 2700001 2810000
WP AAI99683_28 2800001 2910000
WP AAI99683_29 2900001 3010000
WP AAI99683_30 3000001 3110000
WP AAI99683_31 3100001 3210000
WP AAI99683_32 3200001 3310000
WP AAI99683_33 3300001 3410000
WP AAI99683_34 3400001 3510000
WP AAI99683_35 3500001 3610000
WP AAI99683_36 3600001 3710000
WP AAI99683_37 3700001 3810000
WP AAI99683_38 3800001 3910000
WP AAI99683_39 3900001 4010000
WP AAI99683_40 4000001 4110000
WP AAI99683_41 4100001 4210000
WP AAI99683_42 4200001 4310000
WP AAI99683_43 4300001 4403765

Query Match 100.0%; Score 28; DB 4; Length 110000;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACACAGTACGACTC 28
Db 72742 GCTGTGGCGCAGCTACACAGTACGACTC 72769

```

```

RESULT 3
ABLO1952/c
ID ABL01952 standard; cDNA; 5746 BP.
XX AC ABL01952;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 338.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical; gene; ss.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US009231.
XX PR 23-MAR-2000; 2000US-0191637P.
XX PR 11-JUL-2000; 2000US-00614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI; 2001-656860/75.
XX DR P-PSDB; ABB57849.
XX PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signaling and cell-cell
interactions.
XX PS Claim 1; SEQ ID NO 338; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABU16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
ABB72072). The sequence data for this patent did not form part of the
printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 5746 BP; 1776 A; 1219 C; 1123 G; 1628 T; 0 U; 0 Other;

Query Match 65.7%; Score 18.4; DB 4; Length 5746;
Best Local Similarity 78.6%; Pred. No. 1.8e+02;
Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACACAGTACGACTC 28
Db 2598 GCTGTGGCGCAGCTTCACTGTACCGTTC 2571

RESULT 4
ABQ88154/c
ID ABQ88154 standard; cDNA; 147708 BP.
XX AC ABQ88154;
XX DT 18-SEP-2002 (first entry)
XX DE Human osteoblast differentiation related cDNA SEQ ID NO 61.
XX KW Human; osteoblast; stem cell differentiation; bone tissue deposition;
osteoporosis; osteopathic; ss.
XX OS Homo sapiens.

```


XX WO200250301-A2.
 XX 27-JUN-2002.
 XX 18-DEC-2001; 2001WO-US048276.
 XX 18-DEC-2000; 2000US-0255882P.
 XX 24-APR-2001; 2001US-0285691P.
 XX (GENE-) GENE LOGIC INC.
 XX (PROC) PROCTER & GAMBLE CO.
 XX Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A,
 XX Mertz L;
 XX WPI; 2002-557663/59.
 XX Use of genes and their expression profiles associated with osteoblast
 XX differentiation for screening modulators bone formation, for diagnosing
 XX or treating e.g. osteoporosis, or as markers for the differentiation
 XX process.
 XX Claim 1; SEQ ID NO 61; 78pp + Sequence Listing; English.
 XX The invention relates to genes and their expression profiles are used
 XX for: (a) screening modulators of precursor stem cell differentiation into
 XX osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
 XX deposition of bone tissue, abnormal rate of osteoblast formation or
 XX osteoporosis; or (c) treating or monitoring treatment of the conditions
 XX cited in (b), or monitoring the progression of bone tissue deposition.
 XX Specific conditions include postmenopausal osteoporosis, glucocorticoid
 XX osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-
 XX induced abnormalities in bone formation or bone loss, conditions that
 XX involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
 XX skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
 XX or fibrous dysplasia. The present sequence is that of an osteoblast
 XX differentiation associated cDNA marker of the invention. Note: The
 XX sequence data for this patent did not form part of the printed
 XX specification, but was obtained in electronic format directly from WIPO
 XX at ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 147708 BP; 40695 A; 32946 C; 33019 G; 41048 T; 0 U; 0 Other;
 Query Match 65.7%; Score 18.4; DB 6; Length 147708;
 Best Local Similarity 78.6%; Pred. No. 2.6e+02;
 Matches 22; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 QY 1 GCTGTGGCGGAGCTACACAGTACGACTC 28
 Db 115574 GCTGAGGCTCAGCTACACAGCAAGGTTTC 115547
 RESULT 5
 ID ABL79024 standard; cDNA; 356 BP.
 XX ABL79024;
 XX 17-MAY-2002 (first entry)
 XX Human ovarian cancer related cDNA clone SEQ ID NO:2002.
 XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
 XX Homo sapiens.
 XX WO200192581-A2.
 XX 06-DEC-2001.
 XX 29-MAY-2001; 2001WO-US017756.

PR 26-MAY-2000; 2000US-0207484P.
 XX (CORI-) CORIXA CORP.
 XX Algate PA, Harlocker SL, Jones R;
 XX WPI; 2002-122075/16.
 XX Composition for therapy and diagnosis of ovarian cancer comprising
 XX polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
 XX polypeptide, antibody specific to polypeptide or T cell expressing
 XX polypeptide.
 XX Claim 1; SEQ ID NO 2002; 489pp; English.
 XX The present invention describes a composition (I) comprising: carriers
 XX and immunostimulants; and a polypeptide (II) of a ovarian tumour
 XX polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
 XX from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,
 XX (III) encoding (II) having a sequence (S2), a T cell population of (II),
 XX or antigen presenting cells that express (II). (I) has cytostatic
 XX activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
 XX detecting ovarian cancer in a patient's biological sample preferably
 XX serum or ovarian tissue. The method comprises contacting a biological
 XX sample from a patient with (IV), detecting the amount of polynucleotide
 XX hybridising to (IV) and comparing the amount to a predetermined cutoff
 XX value and thereby detecting ovarian cancer in the patient, where the
 XX amount of polynucleotide hybridising to (IV) is detected preferably by
 XX polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
 XX useful for stimulating and/or expanding T cells specific for an ovarian
 XX tumour protein comprising and/or expanding T cells with (III) or (II). (III) is
 XX useful in design and preparation of ribozyme molecules for inhibiting
 XX expression of the tumour polypeptides and proteins in tumour cells; and
 XX to isolate a full length gene from a suitable library e.g., a tumour cDNA
 XX library using well known techniques
 XX SQ Sequence 356 BP; 92 A; 83 C; 70 G; 105 T; 0 U; 6 Other;
 Query Match 64.3%; Score 18; DB 6; Length 356;
 Best Local Similarity 80.8%; Pred. No. 1.9e+02;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 CTGTGGCGGAGCTACACAGTACGACT 27
 Db 8 CTGTGGCTCAGTAAGTAGTAGTACGACT 33
 RESULT 6
 ID AAI93288/c
 XX AAI93288 standard; cDNA; 386 BP.
 XX AAI93288;
 XX 06-NOV-2001 (first entry)
 XX Human polynucleotide SEQ ID NO 13348.
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 XX tissue growth factor; immunomodulatory; cancer; leukaemia;
 XX nervous system disorders; arthritis; inflammation; ss.
 XX Homo sapiens.
 XX WO200164835-A2.
 XX 07-SEP-2001.
 XX 26-FEB-2001; 2001WO-US004927.
 XX 28-FEB-2000; 2000US-00515126.
 XX 18-MAY-2000; 2000US-00577409.

```
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO13357.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
XX
PS Claim 1; SEQ ID NO 13348; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 386 BP; 84 A; 98 C; 83 G; 88 T; 0 U; 33 Other;
Query Match 64.3%; Score 18; DB 4; Length 386;
Best Local Similarity 80.8%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 CTGTGGCGCAGCTACACAGTACGACT 27
DB 75 CTGTGGTGCAGCTACTCGGGAGGACT 50
XX
RESULT 7
AAI85519
ID AAI85519 standard; cDNA; 424 BP.
XX
AC AAI85519;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5579.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
FN WO200164835-A2.
XX
PI 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US004927.
XX
XX 28-FEB-2000; 2000US-00515126.
XX
XX 18-MAY-2000; 2000US-00577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR P-PSDB; AAO05588.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
XX
PS Claim 1; SEQ ID NO 13348; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 424 BP; 125 A; 102 C; 102 G; 95 T; 0 U; 0 Other;
Query Match 64.3%; Score 18; DB 4; Length 424;
Best Local Similarity 80.8%; Pred. No. 2e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 GCTGTGGCGCAGCTACACAGTACGAC 26
DB 227 GATGTGGCGCAACTTCTCAGTATGAC 252
XX
RESULT 8
ACH21700/c
ID ACH21700 standard; cDNA; 473 BP.
XX
AC ACH21700;
XX
DT 13-OCT-2003 (first entry)
XX
DE Human adult ovary cDNA #80..
XX
XX Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW genome mapping; biodiversity; genetic disorder.
XX
OS Homo sapiens.
XX
FN US2003073623-A1.
XX
PD 17-APR-2003.
XX
XX 30-JUL-2001; 2001US-00918995.
XX
XX 30-JUL-2001; 2001US-00918995.
XX
PA (DRMA/) DRMANAC R T.
PA (LABA/) LABAT I.
PA (STAC/) STACHE-CRAIN B.
PA (DICK/) DICKSON M C.
PA (JONE/) JONES L W.
XX
XX Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
XX WPI; 2003-615964/58.
XX
XX New polynucleotide sequences obtained from various cDNA libraries, useful
PT as hybridization probes, as oligomers for PCR, for chromosome and gene
PT mapping, in the recombinant production of protein, or in generating
PT antisense DNA or RNA.
XX
PS Claim 1; SEQ ID NO 8912; 44pp; English.
XX
XX The invention relates to an isolated polynucleotide comprising any one of
CC 38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC determined by the technique of SBH (sequencing by hybridisation). Also
CC included is a purified polypeptide comprising a sequence corresponding to
CC a reading frame of the novel polynucleotide. The nucleic acid sequences
CC are useful in diagnostics as expressed sequence tags (EST) for
```

CC identifying expressed genes or for physical mapping of the human genome, in forensics, in assessing biodiversity, or in identifying mutations responsible for genetic disorders and other traits. The nucleotide sequences are also useful as hybridisation probes, as oligomers for PCR, for chromosome and gene mapping, in the recombinant production of protein, or in generating antisense DNA or RNA. The purified polypeptide is useful for generating antibodies specific for it. The present sequence is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?docID=20030073623

CC Sequence 473 BP; 139 A; 106 C; 110 G; 114 T; 0 U; 4 Other; Query Match 64.3%; Score 18; DB 8; Length 473; Best Local Similarity 80.8%; Pred. No. 2e+02; Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 CTGTGGCGCAGCTACACAGTACGACT 27

DB 330 CTGTGGCTCAGTAAGATAGTACGACT 305

RESULT 9

AAZ23083/c

CC AAZ23083 standard; DNA; 610 BP.

XX AC AAZ23083;

XX 17-JAN-2000 (first entry)

XX 3' UTR sequence of RB binding protein (rbp-2) gene.

XX Human; cancer-associated polypeptide; plu-1; ovarian cancer; vaccine; breast cancer; immune response; cytotoxic T lymphocyte; imaging agent; therapeutic; plu-1 antigen; breast tumor; tumor antigen; humx169a; RB binding protein; rbp-2; hsmcy; UTR; untranslated region; ss.

XX Homo sapiens.

XX W09949034-A1.

XX 30-SEP-1999.

XX 19-MAR-1999; 99WO-GB000866.

XX 20-MAR-1998; 98GB-00005977.

XX (IMCR) IMPERIAL CANCER RES TECHNOLOGY.

XX Taylor-Papadimitriou J;

XX WPI; 1999-591090/50.

XX New nucleic acid encoding the cancer-associated polypeptide plu-1, for diagnosis, treatment and prevention of cancer, especially of breast and ovary.

XX Example 1; Fig 5; 173pp; English.

XX The invention relates to a human cancer-associated polypeptide plu-1. The plu-1 polypeptide can be recombinantly expressed by standard recombinant methodology. Detection of the plu-1 nucleic acid or the polypeptide is used for the following: (i) diagnosis (including imaging) and prognosis of, and determination of susceptibility to, cancer, specifically ovarian or breast cancer; and (ii) treating cancer (by inducing an immune response against cancer cells, e.g. as a vaccine, or by antisense inhibition).

XX Antigens derived from the polypeptide are used to generate activated cytotoxic T lymphocytes, or dendritic cells, for subsequent return to the patient for treatment of cancer. The polypeptide may also be used to identify inhibitors of plu-1 activity. Fragments of the polypeptide, and antibodies raised against plu-1, are useful as assay and imaging agents, also therapeutically (to induce an anti-idiotypic response or where

CC conjugated to cytotoxic agents). The plu-1 antigen is expressed more commonly in breast tumors than some known tumor antigens. Sequences AAZ23080-85 represent 5' UTRs and 3' UTRs of humx169a, rbp-2, hsmcy genes. These sequences are homologous to the 5' and 3' UTRs of the plu-1 gene

CC Sequence 610 BP; 177 A; 125 C; 141 G; 167 T; 0 U; 0 Other; Query Match 64.3%; Score 18; DB 2; Length 610; Best Local Similarity 80.8%; Pred. No. 2.1e+02; Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 CTGTGGCGCAGCTACACAGTACGACT 27

DB 349 CTGTGGCTCAGTAAGATAGTACGACT 324

RESULT 10

ABQ90232/c

CC ABQ90232 standard; DNA; 1014 BP.

XX AC ABQ90232;

XX 01-OCT-2002 (first entry)

XX M. capsulatus gene #217 for DNA array.

XX Micro array; gene; ds; differential expression; gene expression.

XX Methylcoccus capsulatus.

XX W0200255655-A2.

XX 18-JUL-2002.

XX 14-JAN-2002; 2002WO-NO000019.

XX 12-JAN-2001; 2001NO-00000235.

XX 12-JAN-2001; 2001NO-00000239.

XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.

XX (TIGR-) TIGR.

XX Birkeland NK, Eidhammer I, Jonassen I, Jensen KB, Lien T; Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS; Salzberg SL;

XX WPI; 2002-557818/59.

XX Novel DNA array useful for determining differential expression of Methylcoccus capsulatus genes, comprises polynucleotides or oligonucleotides representative for a selective number of Methylcoccus capsulatus genes.

XX Claim 19; Page 156; 678pp; English.

XX The invention relates to a novel DNA array giving a representation of a number of Methylcoccus capsulatus genes. The method of the invention is useful for determination of the differential expression of the genes of M. capsulatus, and for studying gene expression on a genomic scale and in gene expression assays of M. capsulatus genes. The sequences shown in ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the invention

XX Sequence 1014 BP; 217 A; 328 C; 304 G; 165 T; 0 U; 0 Other; Query Match 64.3%; Score 18; DB 6; Length 1014; Best Local Similarity 80.8%; Pred. No. 2.2e+02; Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GCTGTGGCGCAGCTACACAGTACGACT 26

DB 611 GCGGTGGCGGCCACACAGGACGGC 586

RESULT 11
ABAI8457/c
ID ABAI8457 standard; DNA; 2215 BP.
XX
XX
AC ABAI8457;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SRQ ID NO 10788.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antineoplastic; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antischistosomal; antianemic; antithrombotic; cancer;
KW antirheumatic; hepatoprotective; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
FN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PE 17-JAN-2001; 2001WO-US001334.
XX
PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225256P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227005P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 13-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 20-OCT-2000; 2000US-0242221P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246603P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.

PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 01-DEC-2000; 2000US-0251160P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 05-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 11-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Rosen CA, Barash SC, Ruben SM;
 PI
 XX WPI; 2001-541565/60.
 DR
 XX
 XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system cancers
 PT and metastases.
 PT
 XX
 XX Disclosure; SEQ ID NO 10788; 1701pp + Sequence Listing; English.
 PS
 XX
 CC The invention relates to novel genes (ABAI1004-ABAI21534) and proteins
 CC (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
 CC infectious diseases such as viral, bacterial, fungal and parasitic
 CC infections. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 2215 BP; 598 A; 411 C; 474 G; 732 T; 0 U; 0 Other;

 Query Match 64.3%; Score 18; DB 5; Length 2215;
 Best Local Similarity 80.8%; Fred. No. 2.4e+02;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

 QY 2 CTGTGGCGCAGCTACAGTACAGCT 27
 Db 1142 CTGTGGCTCAGTAGATAGTACGACT 1117

 RESULT 12
 ABAI8458/c
 ID ABAI8458 standard; DNA; 2216 BP.
 XX
 XX ABAI8458;
 AC
 XX
 XX 23-JAN-2002 (first entry)
 DT
 XX
 XX Human nervous system related polynucleotide SEQ ID NO 10789.
 DE
 XX Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
 KW antiparkinsonian; antischlicking; antianaemic; antiarthritic; cancer;

KW antirheumatic; hepatotropic; cerebrotropic; antiinflammatory;
 KW anti-allergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
 XX neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
 OS Homo sapiens.
 XX WO200159063-A2.
 PN 16-AUG-2001.
 PD
 XX 17-JAN-2001; 2001WO-US001334.
 FF
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0198874P.
 PR 17-MAR-2000; 2000US-0199076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 11-JUL-2000; 2000US-0217496P.
 PR 14-JUL-2000; 2000US-0218290P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 26-JUL-2000; 2000US-0220964P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226581P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.

XX (GENE-) GENE LOGIC INC.
XX Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX WPI; 2002-435328/46.
XX
XX Detecting granulocyte activation by detecting differential expression of
XX PT genes associated with granulocyte activation, which serves as diagnostic
XX PT markers that is useful for monitoring disease states and drug toxicity.
XX
XX Claim 1; SEQ ID NO 722; 114pp; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
XX CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
XX CC DNA chip analysis as given in the specification, and comparing the
XX CC expression level to an expression level in an unactivated GC where
XX CC differential expression of Gs is indicative of GCA. Also included are
XX CC modulating (M2) GA by contacting GC with an agent that alters the
XX CC expression of at least one gene in Gs; (2) screening (M3) for an agent
XX CC capable of modulating GCA or an inflammation (especially chronic) in a
XX CC tissue, an allergic response in a subject, exposure of a subject to a
XX CC pathogen or sterile inflammatory disease using the gene expression
XX CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
XX CC tissue, an allergic response in a subject, exposure of a subject to a
XX CC pathogen or sterile inflammatory disease, by detecting the level of
XX CC expression in a sample of the tissue of gene(s) from Gs, where the level
XX CC of expression of the gene is indicative of inflammation; (4) treating
XX CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
XX CC response in a subject, exposure of a subject to a pathogen or sterile
XX CC inflammatory disease, by contacting a tissue having inflammation with an
XX CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
XX CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
XX CC for screening an agent capable of modulating GCA preferably in an
XX CC inflammation in a tissue; M4 is useful for detecting an inflammation
XX CC (especially chronic) in a tissue, an allergic response in a subject,
XX CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
XX CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
XX CC cardiac reperfusion injury, renal reperfusion injury, AKDS, adult
XX CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
XX CC disease, ulcerative colitis, periodontal disease; also bacterial
XX CC infection, viral infection, parasitic infection, protozoal infection,
XX CC fungal infection and M5 is useful for treating one of the above
XX CC conditions. The present sequence represents a gene differentially
XX CC expressed in granulocytes. Note: The sequence data for this patent did
XX CC not form part of the printed specification, but was obtained in
XX CC electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 6455 BP; 1868 A; 1311 C; 1672 G; 1604 T; 0 U; 0 Other;

Query Match 64.3%; Score 18; DB 6; Length 6455;
Best Local Similarity 80.8%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTGTGGCGCAGCTACACAGTACGACT 27
DB 6187 CTGTGGCTCAGTAGATAGTACGACT 6162

RESULT 14
ACF79923/c
ID ACF79923 standard; CDNA; 6455 BP.
XX
XX AC ACF79923;
XX
XX 15-JAN-2004 (first entry)
XX
XX Breast cancer specific gene RBBP2 over-expressed in breast cancer.
XX
XX Breast cancer; marker; human; retinoblastoma binding protein 2;
XX KW diagnosis; cytostatic; biochip; vaccine; gene; ss.
XX

OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 153..5321
XX FT /*tag=a
XX FT /*product= "Retinoblastoma binding protein 2"
XX
XX WO2003073911-A2.
XX
XX PD 12-SEP-2003.
XX
XX PF 27-FEB-2003; 2003WO-US005984.
XX
XX PR 28-FEB-2002; 2002US-0359999P.
XX (GEO) UNIV GEORGETOWN.
XX
XX Su YA, Yang J;
XX
XX WPI; 2003-721995/68.
XX P-PSDB; ABM78943.
XX
XX Detecting breast cancer in a subject comprises contacting a biological
XX PT sample with an agent that binds to a polynucleotide or polypeptide of a
XX PT breast-cancer specific gene (BCSG).
XX
XX Claim 6; Page 71-75; 143pp; English.
XX
XX The present sequence is that of a breast cancer specific gene (BCSG)
XX CC identified by microarray gene expression analysis as being over-expressed
XX CC in breast cancers in comparison to healthy tissue. The gene was
XX CC identified as RBBP2 encoding retinoblastoma binding protein 2, which was
XX CC previously known to bind to the tumour suppressor gene RB and to reverse
XX CC RB-mediated suppression of the activity of E2F transcription factor. It is
XX CC one of 19 (see ACF79921-39) BCSGs of the invention that are
XX CC differentially expressed in breast cancer cell lines and breast cancer
XX CC tissue samples as compared to control cell lines and normal tissue
XX CC samples. The polynucleotides transcribed from the BCSGs and the
XX CC polypeptides encoded by the BCSGs are designated as breast cancer
XX CC specific markers (BCSM). The invention provides a method for detection of
XX CC breast cancer by measuring expression levels of BCSGs, and in particular
XX CC the level of polynucleotides transcribed from and polypeptides encoded by
XX CC the BCSGs. A pharmaceutical composition for the treatment of breast
XX CC cancer comprises a BCSM, an antibody directed against a BCSM, a vaccine
XX CC generated using a BCSM, or an agent that modulates an expression level of
XX CC a BCSG or an activity of a BCSM. A biochip for diagnosing breast cancer
XX CC or screening agents that inhibit breast cancer comprises a BCSG or BCSM
XX
XX SQ Sequence 6455 BP; 1868 A; 1311 C; 1672 G; 1604 T; 0 U; 0 Other;

Query Match 64.3%; Score 18; DB 9; Length 6455;
Best Local Similarity 80.8%; Pred. No. 2.7e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTGTGGCGCAGCTACACAGTACGACT 27
DB 6187 CTGTGGCTCAGTAGATAGTACGACT 6162

RESULT 15
ADE62448/c
ID ADE62448 standard; DNA; 6915 BP.
XX
XX AC ADE62448;
XX
XX 29-JAN-2004 (first entry)
XX
XX Human gene AJ243212, SEQ ID NO 8377.
XX
XX Human; ds; gene; pain; neuronal tissue; gene therapy;
XX KW spinal segmental nerve injury; chronic constriction injury; CCI;
XX KW spared nerve injury; SNI; Chung.
XX

OS Homo sapiens.
XX WO2003016475-A2.
XX
XX
PD 27-FEB-2003.
XX
XX 14-AUG-2002; 2002WO-US025765.
XX
XX 14-AUG-2001; 2001US-0312147P.
PR 01-NOV-2001; 2001US-0346382P.
PR 26-NOV-2001; 2001US-0333347P.
XX
XX (GEHO) GEN HOSPITAL CORP.
PA (FARB) BAYER AG.
XX
XX Woolf C, D'urso D, Befort K, Costigan M;
XX WPI; 2003-268312/26.
DR GENBANK; AJ243212.
XX
XX New composition comprising two or more isolated polypeptides, useful for
PT preparing a medicament for treating pain in an animal.
XX
XX Claim 1; Page; 1017pp; English.
XX
XX The invention discloses a composition comprising two or more isolated rat
CC or human polynucleotides or a polynucleotide which represents a fragment,
CC derivative or allelic variation of the nucleic acid sequence. Also
CC claimed are a vector comprising the novel polynucleotide, a host cell
CC comprising the vector, a method for identifying a nucleotide sequence
CC which is differentially regulated in an animal subjected to pain and a
CC kit to perform the method, an array, a method for identifying an agent
CC that increases or decreases the expression of the polynucleotide sequence
CC that is differentially expressed in neuronal tissue of a first animal
CC subjected to pain, a method for identifying a compound which regulates
CC the expression of a polynucleotide sequence which is differentially
CC expressed in an animal subjected to pain, a method for identifying a
CC compound that regulates the activity of one or more of the
CC polynucleotides, a method for producing a pharmaceutical composition, a
CC method for identifying a compound or small molecule that regulates the
CC activity in an animal of one or more of the polypeptides given in the
CC specification, a method for identifying a compound useful in treating
CC pain and a pharmaceutical composition comprising the one or more
CC polypeptides or their antibodies. The polynucleotide or the compound that
CC modulates its activity is useful for preparing a medicament for treating
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. Gene
CC therapy). The sequence presented is a human DNA (shown in Table 2 of the
CC specification) which encodes one of the polypeptides of the invention
CC which is differentially expressed during pain. Note: The sequence data
CC for this patent did not form part of the printed specification, but was
CC obtained in electronic form directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6915 BP; 1462 A; 1915 C; 1972 G; 1566 T; 0 U; 0 Other;

Query Match 64.3%; Score 18; DB 9; Length 5915;
Best Local Similarity 80.8%; Pred. No. 2.8e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 TGTGGCGCAGCTACACAGTACGACTC 28
||||| ||||| ||||| ||||| |||||
Db 2997 TGTGGGGCGCTCCACAGTACGACTC 2972

Search completed: April 29, 2004, 04:57:00
Job time : 234.237 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:01:30 ; Search time 2207.83 Seconds
(without alignments)
378.717 Million cell updates/sec

Title: US-10-624-714-5
Perfect score: 28
Sequence: 1 gctgtggcgacgtacacagtcagctc 28

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:**

- 1: em_estba:**
- 2: em_estbm:**
- 3: em_estbn:**
- 4: em_estbv:**
- 5: em_estbv:**
- 6: em_estbv:**
- 7: em_estbv:**
- 8: em_estbv:**
- 9: gb_est1:**
- 10: gb_est2:**
- 11: gb_est3:**
- 12: gb_est3:**
- 13: gb_est4:**
- 14: gb_est5:**
- 15: em_estfun:**
- 16: em_estfun:**
- 17: em_gss_hum:**
- 18: em_gss_inv:**
- 19: em_gss_pln:**
- 20: em_gss_vrt:**
- 21: em_gss_fun:**
- 22: em_gss_mam:**
- 23: em_gss_mus:**
- 24: em_gss_pro:**
- 25: em_gss_rod:**
- 26: em_gss_phg:**
- 27: em_gss_vrl:**
- 28: gb_gss1:**
- 29: gb_gss2:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	21.6	77.1	551	13	BU406463
2	21.6	77.1	687	14	CA786287
3	21.6	77.1	707	13	BU256022
4	21.6	77.1	768	13	BU355788

5	21.6	77.1	771	13	BU336057
6	21.6	77.1	793	13	BU276298
7	21.6	77.1	825	14	CF254672
8	21.6	77.1	827	13	BU363287
9	21.6	77.1	849	13	BU112074
10	21.6	77.1	863	13	BU450608
11	21.6	77.1	894	13	BU331214
12	19.6	70.0	581	29	CE193727
13	19.6	70.0	667	29	EX121035
14	19.2	68.6	303	10	BB098794
15	19.2	68.6	764	29	CG439762
16	19.2	68.6	802	29	CG358430
17	19.2	68.6	822	29	CG438762
18	19.2	68.6	833	29	CG845566
19	19.2	68.6	843	29	CG264886
20	19.2	68.6	910	29	CG439770
21	19.2	68.6	920	29	CG052248
22	19.2	68.6	930	29	AG210106
23	19.2	68.6	940	29	AG207249
24	19.2	68.6	940	29	CE566816
25	19.2	68.6	940	29	AG209812
26	19.2	68.6	940	29	CA163676
27	19.2	68.6	940	29	AG210446
28	19.2	68.6	940	29	AG204016
29	19.2	68.6	940	29	AG206008
30	19.2	68.6	940	29	CA144501
31	19.2	68.6	940	29	AG207305
32	19.2	68.6	940	29	CA156311
33	19.2	68.6	940	29	AG207112
34	19.2	68.6	940	29	AG205528
35	19.2	68.6	940	29	AG204041
36	19.2	68.6	940	29	AG202306
37	19.2	68.6	940	29	AG205611
38	19.2	68.6	940	29	CG162966
39	19.2	68.6	940	29	AG170294
40	18.8	67.1	518	10	BE554128
41	18.8	67.1	549	13	BX511465
42	18.8	67.1	556	9	AA866721
43	18.8	67.1	671	10	BB267243
44	18.8	67.1	683	12	BU270921
45	18.8	67.1	684	12	BG093370

ALIGNMENTS

RESULT 1
BU406463

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

BU406463 604138238F1 CSEQCHN59 Gallus gallus cdna clone CHEST96619 5', mRNA sequence.
BU406463 604138238F1 CSEQCHN59 Gallus gallus cdna clone CHEST96619 5', mRNA sequence.
BU406463 604138238F1 CSEQCHN59 Gallus gallus cdna clone CHEST96619 5', mRNA sequence.
EST. GI:25775519
Gallus gallus (chicken)
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 551)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445392
Contact: Simon Hubbard
Department of Biomedical Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409

BU336057 603869788
BU276298 603600417
CF254672 mdvnl19_b
BU363287 603584913
BU112074 603126183
BU450608 603216442
BU331214 603497917
CE193727 tigr-988-
EX121035 dario rer
BB098794 BB098794
CG439762 OGVHP82TH
CG358430 OGVHP82TH
CG438762 OGVHP82TH
CG845566 OGVHP82TH
CG264886 OGVHP82TH
CG439770 OGVHP82TH
CG052248 PUFRES8TB
AG210106 Oryza sat
AG207249 Oryza sat
CE566816 tigr-988-
AG209812 Oryza sat
CA163676 SCRUR308
AG210446 Oryza sat
AG204016 Oryza sat
AG206008 Oryza sat
CA144501 SCSPR206
AG207305 Oryza sat
CA156311 SCSPR206
AG207112 Oryza sat
AG205528 Oryza sat
AG204041 Oryza sat
AG202306 Oryza sat
AG205611 Oryza sat
CG162966 PUKBB2TB
AG170294 Pan trogl
BE554128 ur38e03.Y
BX511465 BX511465
AA866721 vx85c01.r
BB267243 BB267243
BU270921 BJ270921
BG093370 ut70f07.Y

Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source

Location/Qualifiers
1..551
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST35g21"
/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSEQCHN59"

/note="Organ: limbs; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Query Match 77.1%; Score 21.6; DB 13; Length 551;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCTGTGGCGCAGCTACACGATCAGTAC 28
Db 275 GCTGTGGCTCAGCTGCAAGTACGACTC 302

RESULT 2
CA786287
LOCUS

DEFINITION CA786287 687 bp mRNA linear EST 04-DEC-2002
cDNA clone pgnic.pk011.i3 Chicken lymphoid cDNA library (pgnic) Gallus gallus
F-box only protein 5 (Homo sapiens), mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

CA786287 1 GI:26049834
EST.
Gallus gallus (chicken)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 687)
Morgan, R.W. and Burnside, J.
Chicken lymphoid ESTs
Unpublished (2001)
Contact: Robin W. Morgan
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1341
Fax: 302-831-2822
Email: morganudel.edu, www.chickest.udel.edu.

FEATURES
source

Location/Qualifiers
1..687
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="pgnic.pk011.i3"
/sex="Male and Female"
/tissue_type="thymus, bursa, spleen, PBL, bone marrow"
/lab_host="E.coli ENDH10B"
/clone_lib="Chicken lymphoid cDNA library (pgnic)"
/note="vector: pcwmsport 6"

ORIGIN

Query Match 77.1%; Score 21.6; DB 13; Length 707;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCTGTGGCGCAGCTACACGATCAGTAC 28
Db 69 GCTGTGGCTCAGCTGCAAGTACGACTC 96

RESULT 4

Query Match 77.1%; Score 21.6; DB 14; Length 687;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCTGTGGCGCAGCTACACGATCAGTAC 28
Db 549 GCTGTGGCTCAGCTGCAAGTACGACTC 576

RESULT 3
BU256022
LOCUS

DEFINITION BU256022 707 bp mRNA linear EST 26-NOV-2002
603413784F1 CSEQCHN38 Gallus gallus cDNA clone CHEST35g21 5', mRNA
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BU256022 1 GI:25514412
EST.
Gallus gallus (chicken)

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 707)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

22335534
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)

PO Box 88 Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES
source

Location/Qualifiers
1..707
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST35g21"
/dev_stage="36"
/lab_host="DH10B"
/clone_lib="CSEQCHN38"

/note="Organ: limbs; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

BU355788
 LOCUS 603407405F1 CSEQCHN70 768 bp mRNA linear EST 28-NOV-2002
 DEFINITION Gallus gallus CDNA clone CHEST31812 5', mRNA
 sequence.
 ACCESSION BU355788
 VERSION BU355788.1 GI:25863789
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 768)
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 FEATURES
 source
 1..768
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST31812"
 /dev_stage="36"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN70"
 /note="Organ: hearts; Vector: pBluescript II KS(+);
 Site 1: EcoRI; Site 2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunted, ligated to NotI adapters, digested with
 EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN

Query Match 77.1%; Score 21.6; DB 13; Length 768;
 Best Local Similarity 85.7%; Pred. NO. 1.8e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 GCTGTGGCGCAGCTACACAGTACGACTC 28
 Db 609 GCTGTGGTCCACCTGCAAGTACGACTC 636

ORIGIN

Query Match 77.1%; Score 21.6; DB 13; Length 768;
 Best Local Similarity 85.7%; Pred. NO. 1.8e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 GCTGTGGCGCAGCTACACAGTACGACTC 28
 Db 609 GCTGTGGTCCACCTGCAAGTACGACTC 636

RESULT 5
 BU336057
 LOCUS 603869788F1 CSEQCHN65 771 bp mRNA linear EST 28-NOV-2002
 DEFINITION Gallus gallus CDNA clone CHEST394022 5', mRNA
 sequence.
 ACCESSION BU336057
 VERSION BU336057.1 GI:25844058
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 771)
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392
 COMMENT Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.
 FEATURES
 source
 1..771
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST894022"
 /tissue_type="whole embryo"
 /dev_stage="10"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN65"
 /note="Organ: whole embryo; Vector: pBluescript II KS(+);
 Site 1: EcoRI; Site 2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunted, ligated to NotI adapters, digested with
 EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bernaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN

Query Match 77.1%; Score 21.6; DB 13; Length 771;
 Best Local Similarity 85.7%; Pred. NO. 1.8e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 GCTGTGGCGCAGCTACACAGTACGACTC 28
 Db 40 GCTGTGGCTCACCTGCAAGTACGACTC 67

ORIGIN

Query Match 77.1%; Score 21.6; DB 13; Length 771;
 Best Local Similarity 85.7%; Pred. NO. 1.8e+02;
 Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Qy 1 GCTGTGGCGCAGCTACACAGTACGACTC 28
 Db 40 GCTGTGGCTCACCTGCAAGTACGACTC 67

ORIGIN

RESULT 6
 BU276298
 LOCUS 603600417F1 CSEQCHN54 793 bp mRNA linear EST 27-NOV-2002
 DEFINITION Gallus gallus CDNA clone CHEST574a6 5', mRNA
 sequence.
 ACCESSION BU276298
 VERSION BU276298.1 GI:25725752
 KEYWORDS EST.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

ORIGIN

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 793)
 AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 TITLE A Comprehensive Collection of Chicken cDNAs
 JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
 MEDLINE 22335534
 PUBMED 12445392

COMMENT

Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source

1..793
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="ChEST57446"
/sex="Female"
/tissue_type="not cerebrum or cerebellum"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHN54"
/note="Organ: brain; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996) 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Query Match 77.1%; Score 21.6; DB 13; Length 793;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGTACACAGTACGACTC 28
|||||
DB 134 GCTGTGGCTCACCTGCAAGTACGACTC 161
|||||

RESULT 7

CF254672
LOCUS CF254672 825 bp mRNA linear EST 07-AUG-2003
DEFINITION mvn119_b01_r1 Marek's disease virus-infected spleen Gallus gallus
cDNA, mRNA sequence.
CF254672
VERSION CF254672.1 GI:33487927
KEYWORDS EST.
SOURCE Gallus gallus (chicken)

ORGANISM

Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 825)

REFERENCE

AUTHORS Wittzell,H., Bed'Hom,B., Morin,V., Young,J.R., Whittaker,C.J.,
Chause,A.M. and Zoorob,R.
TITLE A collection of chicken ESTs from activated immune cells
JOURNAL Unpublished (2003)

COMMENT

Contact: Zoorob R
UPR 1983
CNRS
7 rue Guy Moquet, BP 8, 94801 Villejuif cedex, France
Tel: 33 1 49 58 35 00
Fax: 33 1 49 58 33 81
Email: zoorob@vjf.cnrs.fr.
Location/Qualifiers
1..825
/organism="Gallus gallus"

FEATURES

source

1..825
/organism="Gallus gallus"

ORIGIN

Query Match 77.1%; Score 21.6; DB 14; Length 825;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGTACACAGTACGACTC 28
|||||
DB 142 GCTGTGGCTCACCTGCAAGTACGACTC 169
|||||

RESULT 8

BU363287 827 bp mRNA linear EST 28-NOV-2002
LOCUS BU363287 827 bp mRNA linear EST 28-NOV-2002
DEFINITION 603584913F1 CSEQCHN72 Gallus gallus cDNA clone ChEST539j4 5', mRNA
sequence.
BU363287
VERSION BU363287.1 GI:25871288
KEYWORDS EST.
SOURCE Gallus gallus (chicken)

ORGANISM

Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 827)

REFERENCE

AUTHORS Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E.,
Fong,W.T., Tickle,C., Brown,W.R.A., Wilson,S.A. and Hubbard,S.J.
TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL A Comprehensive Collection of Chicken cDNAs
MEDLINE Curr. Biol. 12 (22), 1965-1969 (2002)
22355534
PUBMED 12445392

COMMENT

Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source

1..827
Location/Qualifiers
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="ChEST539j4"
/sex="Female"
/tissue_type="cerebrum"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSEQCHN72"

/note="Organ: brain; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunted, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996) 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Query Match 77.1%; Score 21.6; DB 13; Length 827;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;

TITLE A Comprehensive Collection of Chicken cDNAs
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 2235534
PUBMED 12445392
COMMENT

Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

FEATURES

source

1. .894
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHEST410b6"
/tissue_type="whole embryo"
/dev_stage="10"
/lab_host="DH10B"
/clone_lib="CSECHN65"

/note="Organ: whole embryo; Vector: pBluescript II KS(+); Site 1: EcoRI; Site 2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN

Query Match 77.1%; Score 21.6; DB 13; Length 894;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCTGTGGCGGAGCTACACAGTACGACTC 28
Db 119 GCTGTGGCTCACCTGCAAGTACGACTC 146

RESULT 12

CE193727
LOCUS tigr-gss-dog-17000371482974 Dog Library Canis familiaris genomic,
DEFINITION genomic survey sequence.
ACCESSION CE193727
VERSION CE193727.1 GI:35349380
KEYWORDS GSS.
SOURCE Canis familiaris (dog)
ORGANISM Canis familiaris

Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE 1 (bases 1 to 581)
Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.H., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M. and Venter, J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5841), 1898-1903 (2003)

TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA
Tel: 301-838-0200

Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.

FEATURES

source

Location/Qualifiers
1. .581
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/note="Site 1: BatXI; Libraries were prepared from peripheral blood"

ORIGIN

Query Match 70.0%; Score 19.6; DB 29; Length 581;
Best Local Similarity 84.6%; Pred. No. 1.1e+03;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 2 CTGTGGCGGAGCTACACAGTACGACT 27
Db 286 CTGTGGAGGAGCTGCACAGTACCACT 311

RESULT 13

BX121035
LOCUS Danio rerio genomic clone DKEY-160G6, linear GSS 13-MAR-2003
DEFINITION
ACCESSION BX121035
VERSION BX121035.1 GI:27951958
KEYWORDS GSS.
SOURCE Danio rerio (zebrafish)

ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 667)
Humphray, S.J., Huckle, E. and Durham, J.L.
AUTHORS Direct Submission
TITLE Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome
JOURNAL Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humphray@sanger.ac.uk Unpublished
COMMENT This sequence was generated from the T7 end of BAC 160G6. 160G6 is part of the Daniokey BAC Library created by R. Piasterk and N.V. Keygene. Further details:
http://www.sanger.ac.uk/Projects/D_rerio/.

FEATURES

source

Location/Qualifiers
1. .667
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-160G6"
/tissue_type="Testis"
/note="vector pIndigoBAC-536"

ORIGIN

Query Match 70.0%; Score 19.6; DB 29; Length 667;
Best Local Similarity 84.6%; Pred. No. 1.2e+03;
Matches 22; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 TGTGGCGGAGCTACACAGTACGACTC 28
Db 523 TGTGGCGGAGCGCCACAAAGGACTC 548

RESULT 14

BB098794
LOCUS BB098794 RIKEN full-length enriched, 12 days embryo, embryonic body
DEFINITION between diaphragm region and neck Mus musculus cDNA clone
9430065H24 3', mRNA sequence.
ACCESSION BB098794
VERSION BB098794.1 GI:8743828
KEYWORDS EST.

303 bp mRNA

EST 26-JUN-2000

SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 303)
 REFERENCE Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, K., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Tagawa, A., Takahashi, F., Tominaga, N., Toya, T., Tsunoda, Y., Watanabe, A., Watanabe, S., Yamamura, T., Yamana, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
 RIKEN Mouse ESTs (Konno, H., et al.)
 Unpublished (2000)
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Kitsuai, T., Akiyama, Y., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
 Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
 FEATURES source
 1..303
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="9430065H24"
 /tissue_type="embryonic body between diaphragm region and neck"
 /dev_stage="12 days embryo"
 /lab_host="DH10B"
 /clone_lib="RIKEN full-length enriched, 12 days embryo, embryonic body between diaphragm region and neck"
 /note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATTCGAGTTAATTAATTCCTCCCCCCCCC 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 370.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATTCGAGTTAATTAATTCCTCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI."
 ORIGIN

Query Match 68.6%; Score 19.2; DB 10; Length 303;
 Best Local Similarity 87.5%; Pred. No. 1.2e+03;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 5 TGGCGCAGCTACAGTACGACTC 28
 Db 60 TGGCACACTACAGTACTAGC 83
 RESULT 15
 CG439762
 LOCUS CG439762
 DEFINITION CG439762 764 bp DNA linear GSS 17-SEP-2003
 genomic survey sequence.
 ACCESSION CG439762
 VERSION CG439762.1 GI:34816511
 KEYWORDS GSS.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 764)
 REFERENCE Whitelaw, C.A., Quackenbush, J., Van Aken, S., Uterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
 Consortium for Maize Genomics
 Unpublished (2002)
 Other_GSSs: CGVHP82TV
 Contact: Cathy Whitelaw
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-5843
 Fax: 301-838-0208
 Email: whitelaw@tigr.org
 Seq primer: IR
 Class: sheared ends.
 Location/Qualifiers
 1..764
 /organism="Zea mays"
 /mol_type="genomic DNA"
 /strain="B73"
 /db_xref="taxon:4577"
 /clone="ZMMEM0485M20"
 /clone_lib="ZM 0.7 1.5 KB"
 /note="Vector: pBCSK; Site 1: HindII; 0.7-1.5 kb methylation filtered genomic DNA library"
 ORIGIN

Query Match 68.6%; Score 19.2; DB 29; Length 764;
 Best Local Similarity 87.5%; Pred. No. 1.8e+03;
 Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GCTGTGGCGCAGCTACAGTACG 24
 Db 687 GCAGTGGCGAGCTACCCAGTACG 710
 Search completed: April 29, 2004, 11:36:57
 Job time : 2210.83 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 04:13:15 ; Search time 49,4118 Seconds
(without alignments)
314.472 Million cell updates/sec

Title: US-10-624-714-5

Perfect score: 28

Sequence: 1 gctgtggcgcagctacacagctacgactc 28

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	100.0	4403765	3	US-09-103-840A-2
2	28	100.0	4411529	3	US-09-103-840A-1
3	18.2	65.0	1380	4	US-09-252-991A-15862
4	18.2	65.0	1836	4	US-09-252-991A-15831
5	18.2	65.0	2055	4	US-09-252-991A-15771
6	18.2	65.0	2106	4	US-09-252-991A-15891
7	17.4	62.1	3808	2	US-08-916-917-3
8	17.4	62.1	3808	2	US-08-972-631-3
9	17.4	62.1	3808	2	US-08-972-629-3
10	17.4	62.1	3808	2	US-08-972-630-3
11	17.4	62.1	3808	2	US-08-972-211-3
12	17.4	62.1	3808	3	US-09-225-170-3
13	17.4	62.1	4134	2	US-08-817-090B-1
14	17.4	62.1	4137	2	US-08-817-090B-3
15	17.4	62.1	5162	2	US-08-916-917-13
16	17.4	62.1	5162	3	US-09-225-170-13
17	17.2	61.4	14759	4	US-09-661-887-1
18	16.8	60.0	471	4	US-09-465-558-43
19	16.8	60.0	1277	4	US-08-722-570-15
20	16.8	60.0	1277	4	US-08-932-411A-16
21	16.6	59.3	4403765	3	US-09-103-840A-2
22	16.6	59.3	4411529	3	US-09-103-840A-1
23	16.4	58.6	1524	2	US-08-716-942-24
24	16.4	58.6	1524	4	US-09-130-337A-24
25	16.4	58.6	5567	4	US-08-899-241-1
26	16.4	58.6	5943	4	US-09-976-594-272
27	16.4	58.6	148567	4	US-09-801-876B-3

28 16.4 58.6 148567 4 US-10-254-869-3 Sequence 3, Appli
29 16.2 57.9 861 4 US-09-252-991A-14498 Sequence 14498, A
30 16.2 57.9 873 4 US-09-252-991A-14863 Sequence 14863, A
31 16.2 57.9 996 4 US-09-252-991A-15120 Sequence 15120, A
32 16.2 57.9 1029 4 US-09-489-039A-7100 Sequence 7100, Ap
33 16 57.1 157 4 US-09-441-340-16 Sequence 16, Appl
34 16 57.1 432 4 US-09-441-340-19 Sequence 19, Appl
35 16 57.1 474 4 US-09-134-001C-2266 Sequence 2266, Ap
36 16 57.1 1059 4 US-09-489-039A-6329 Sequence 6329, Ap
37 16 57.1 1216 4 US-09-016-434-1474 Sequence 1474, Ap
38 16 57.1 1488 2 US-08-812-203-4 Sequence 4, Appli
39 16 57.1 1488 3 US-09-300-864-4 Sequence 4, Appli
40 16 57.1 1488 4 US-09-598-418-4 Sequence 4, Appli
41 16 57.1 1842 4 US-09-489-039A-6050 Sequence 6050, Ap
42 16 57.1 1901 4 US-09-700-971-1 Sequence 1, Appli
43 16 57.1 2107 4 US-09-441-340-29 Sequence 29, Appl
44 16 57.1 2122 4 US-09-441-340-25 Sequence 25, Appl
45 16 57.1 2378 4 US-09-441-340-27 Sequence 27, Appl

ALIGNMENTS

RESULT 1

US-09-103-840A-2

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

US-09-103-840A-2

Query Match 100.0%; Score 28; DB 3; Length 4403765;

Best Local Similarity 100.0%; Pred. No. 0.0015;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACAGTACGACTC 28

Db 472742 GCTGTGGCGCAGCTACAGTACGACTC 472769

RESULT 2

US-09-103-840A-1

; Sequence 1, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      100.0%; Score 28; DB 3; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 0.0015;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACAGTAC 28
DB 472658 GCTGTGGCGCAGCTACAGTAC 472685

RESULT 3
US-09-252-991A-15862/c
; Sequence 15862, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15862
; LENGTH: 1380
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15862

Query Match      65.0%; Score 18.2; DB 4; Length 1380;
Best Local Similarity 87.0%; Pred. No. 18;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACAGTAC 23
DB 1282 GCTGTGGCGCAGCTTCCCGTGC 1260

RESULT 4
US-09-252-991A-15831
; Sequence 15831, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15831
; LENGTH: 1836
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15831

Query Match      65.0%; Score 18.2; DB 4; Length 1836;
Best Local Similarity 87.0%; Pred. No. 19;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACAGTAC 23
DB 1644 GCTGTGGCGCAGCTTCCCGTGC 1622

RESULT 7
US-08-916-917-3
; Sequence 3, Application US/08916917
; Patent No. 5856132
; GENERAL INFORMATION:
```

APPLICANT: Stephens, Len
APPLICANT: Hawkins, Phillip Thomas
APPLICANT: Braselmann, Sylvia
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3', KINASE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds, LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2811
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,917
FILING DATE: 15-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/672,211
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Abrams, Samuel B
REGISTRATION NUMBER: 30,605
REFERENCE/DOCKET NUMBER: 8549-0006-999
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3808 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-916-917-3

Query Match 62.1%; Score 17.4; DB 2; Length 3808;
Best Local Similarity 77.8%; Pred No. 51;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACAGTACGACT 27
|||||
Db 2256 GCTGTGGCGCAGCCATGCTGCAGACT 2282

RESULT 8
US-08-972-631-3
Sequence 3, Application US/08972631
Patent No. 5856133
GENERAL INFORMATION:
APPLICANT: Stephens, Len
APPLICANT: Hawkins, Phillip T.
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3', KINASE
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/972,631
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/672,211
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 8549-0005-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)854-3660
TELEFAX: (415)854-3694
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3808 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-972-631-3

Query Match 62.1%; Score 17.4; DB 2; Length 3808;
Best Local Similarity 77.8%; Pred No. 51;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACAGTACGACT 27
|||||
Db 2256 GCTGTGGCGCAGCCATGCTGCAGACT 2282

RESULT 9
US-08-972-629-3
Sequence 3, Application US/08972629
Patent No. 5859201
GENERAL INFORMATION:
APPLICANT: Stephens, Len
APPLICANT: Hawkins, Phillip T.
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3', KINASE
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,629
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/672,211
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 8549-0005-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)854-3660
TELEFAX: (415)854-3694
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3808 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown

MOLECULE TYPE: CDNA
US-08-972-629-3
Query Match 62.1%; Score 17.4; DB 2; Length 3808;
Best Local Similarity 77.8%; Pred. No. 51;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 GCTGTGGCGCAGCTACAGTACGACT 27
DB 2256 GCTGTGGCAGCCATGCTGCAGACT 2282
RESULT 10
US-08-972-630-3
Sequence 3, Application US/08972630
Patent No. 5869271
GENERAL INFORMATION:
APPLICANT: Stephens, Len
APPLICANT: Hawkins, Phillip T.
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/972,630
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/672,211
FILING DATE: 27-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 8549-0005-999
TELEPHONE: (415)854-3660
TELEFAX: (415)854-3694
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3808 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
US-08-972-630-3
Query Match 62.1%; Score 17.4; DB 2; Length 3808;
Best Local Similarity 77.8%; Pred. No. 51;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 GCTGTGGCGCAGCTACAGTACGACT 27
DB 2256 GCTGTGGCAGCCATGCTGCAGACT 2282
RESULT 11
US-08-672-211-3
Sequence 3, Application US/08672211
Patent No. 5874273
GENERAL INFORMATION:
APPLICANT: Stephens, Len
APPLICANT: Hawkins, Phillip T.

TITLE OF INVENTION: G-BETA-GAMMA REGULATED
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 2730 Sand Hill Road
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/672,211
FILING DATE: 27-JUN-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Halluin, Albert P.
REGISTRATION NUMBER: 25,277
REFERENCE/DOCKET NUMBER: 8549-0005-999
TELEPHONE: (415)854-3660
TELEFAX: (415)854-3694
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3808 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
US-08-672-211-3
Query Match 62.1%; Score 17.4; DB 2; Length 3808;
Best Local Similarity 77.8%; Pred. No. 51;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 GCTGTGGCGCAGCTACAGTACGACT 27
DB 2256 GCTGTGGCAGCCATGCTGCAGACT 2282
RESULT 12
US-09-225-170-3
Sequence 3, Application US/09225170
Patent No. 6017763
GENERAL INFORMATION:
APPLICANT: Stephens, Len
APPLICANT: Hawkins, Phillip Thomas
APPLICANT: Braselmann, Sylvia
TITLE OF INVENTION: G-BETA-GAMMA REGULATED
TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds, LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2811
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/225,170
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/916,917

;; FILING DATE: 15-AUG-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Abrams, Samuel B
;; REGISTRATION NUMBER: 30,605
;; REFERENCE/DOCKET NUMBER: 8549-0006-999
;; TELEPHONE: 650-493-4935
;; TELEFAX: 650-493-5556
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-09-225-170-3
Query Match 62.1%; Score 17.4; DB 3; Length 3808;
Best Local Similarity 77.8%; Pred. No. 51;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 GCTGTGGCGCAGCTACACAGTACGACT 27
Db 2256 GCTGTGGCAGCATGCTGCACGACT 2282
RESULT 13
US-08-817-090B-1
; Sequence 1, Application US/08817090B
; Patent No. 5885777
; GENERAL INFORMATION:
; APPLICANT: Stoyanov, Borislav
; APPLICANT: Hanck, Theodor
; APPLICANT: Wetzker, Reinhard
; TITLE OF INVENTION: CLONING, EXPRESSION AND CHARACTERIZATION OF
; A NEW FORM OF PHOSPHATIDYLINOSITOL-3-KINASE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,090B
; FILING DATE: 11-APR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 36 696.5
; FILING DATE: 13-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 45 562.3
; FILING DATE: 20-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Betman, Richard J.
; REGISTRATION NUMBER: 39,107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4134 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; QUERY MATCH 62.1%; Score 17.4; DB 3; Length 4137;
Best Local Similarity 77.8%; Pred. No. 52;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

;; ORGANISM: Homo sapiens
;; FEATURE: CDS
;; NAME/KEY: CDS
;; LOCATION: 423..3569
;; US-08-817-090B-1
Query Match 62.1%; Score 17.4; DB 2; Length 4134;
Best Local Similarity 77.8%; Pred. No. 52;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 GCTGTGGCGCAGCTACACAGTACGACT 27
Db 2485 GCTGTGGCAGCATGCTGCACGACT 2511
RESULT 14
US-08-817-090B-3
; Sequence 3, Application US/08817090B
; Patent No. 5885777
; GENERAL INFORMATION:
; APPLICANT: Stoyanov, Borislav
; APPLICANT: Hanck, Theodor
; APPLICANT: Wetzker, Reinhard
; TITLE OF INVENTION: CLONING, EXPRESSION AND CHARACTERIZATION OF
; A NEW FORM OF PHOSPHATIDYLINOSITOL-3-KINASE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nikaido, Marmelstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,090B
; FILING DATE: 11-APR-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 36 696.5
; FILING DATE: 13-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 44 45 562.3
; FILING DATE: 20-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Berman, Richard J.
; REGISTRATION NUMBER: 39,107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)638-5000
; TELEFAX: (202)638-4810
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4137 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 423..3572
;; US-08-817-090B-3
Query Match 62.1%; Score 17.4; DB 2; Length 4137;
Best Local Similarity 77.8%; Pred. No. 52;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 GCTGTGGCGCAGCTACACAGTACGACT 27

DB 2488 GCTGTGGCAGCCATGCTGCAGCACT 2514
|||||
RESULT 15
US-08-916-917-13
; Sequence 13, Application US/08916917
; Patent No. 5856132
; GENERAL INFORMATION:
; APPLICANT: Stephens, Len
; APPLICANT: Hawkins, Phillip Thomas
; APPLICANT: Braselmann, Sylvia
; TITLE OF INVENTION: G-BETA-GAMMA REGULATED
; TITLE OF INVENTION: PHOSPHATIDYLINOSITOL-3' KINASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,917
; FILING DATE: 15-AUG-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/672,211
; FILING DATE: 27-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Abrams, Samuel B
; REGISTRATION NUMBER: 30,605
; REFERENCE/DOCKET NUMBER: 8549-0006-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5162 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-916-917-13

Query Match 62.1%; Score 17.4; DB 2; Length 5162;
Best Local Similarity 77.8%; Pred. No. 54;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 GCTGTGGCGGAGTACAGTACAGTACACT 27
DB 2251 GCTGTGGCAGCCATGCTGCAGCACT 2277
|||||

Search completed: April 29, 2004, 11:45:45
Job time : 65.4118 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 06:08:26 ; Search time 522.667 Seconds

Title: US-10-624-714-5
Perfect score: 28
Sequence: 1 gctgtggcagctacacagtacgactc 28

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	18	64.3	356	9	US-09-867-701-2002
2	18	64.3	473	10	US-09-918-995-8912
3	18	64.3	6455	16	US-10-373-801-3
4	17.6	62.9	1974	15	US-10-156-761-2174
5	17.6	62.9	4671	16	US-10-369-493-25911
6	17.6	62.9	9025608	15	US-10-156-761-1
7	17.4	62.1	627	15	US-10-101-235A-1
8	17.4	62.1	1323	16	US-10-369-493-43781
9	17.4	62.1	1704	13	US-10-282-122A-33322
10	17.4	62.1	2000	16	US-10-260-238-2607
11	17.4	62.1	2726	13	US-09-954-456-1116
12	17.4	62.1	2726	13	US-10-342-887-362
13	17.4	62.1	2726	13	US-10-172-118-382
14	17.4	62.1	2726	13	US-10-171-581-137

c 15	17.4	62.1	3079	15	US-10-084-817-313	Sequence 313, App
c 16	17.4	62.1	3082	15	US-10-240-965-222	Sequence 222, App
c 17	17.4	62.1	3147	15	US-10-119-926-49	Sequence 49, Appli
c 18	17.4	62.1	3237	15	US-10-101-235A-5	Sequence 5, Appli
c 19	17.4	62.1	3342	15	US-10-101-235A-3	Sequence 3, Appli
c 20	17.4	62.1	3649	9	US-09-801-574-15	Sequence 15, Appli
c 21	17.4	62.1	5309	16	US-10-334-143-120	Sequence 120, App
c 22	17.4	62.1	5310	13	US-10-087-192-1571	Sequence 1571, Ap
c 23	17.4	62.1	61588	13	US-10-087-192-1570	Sequence 1570, Ap
c 24	17.4	62.1	94001	17	US-10-210-838-20	Sequence 20, Appli
c 25	17.2	61.4	60	10	US-09-908-975-14246	Sequence 14246, A
c 26	17.2	61.4	2064	13	US-10-282-122A-13230	Sequence 13230, A
c 27	17.2	61.4	3094	9	US-09-923-297-220	Sequence 220, App
c 28	17.2	61.4	3442	16	US-10-159-563-122	Sequence 122, App
c 29	17.2	61.4	14759	9	US-09-952-360-1	Sequence 1, Appli
c 30	17.2	61.4	14759	15	US-10-346-723-1	Sequence 1, Appli
c 31	17	60.7	554	13	US-10-027-632-322230	Sequence 322230,
c 32	17	60.7	554	13	US-10-027-632-322231	Sequence 322231,
c 33	17	60.7	554	16	US-10-027-632-322230	Sequence 322230,
c 34	17	60.7	554	16	US-10-027-632-322231	Sequence 322231,
c 35	17	60.7	1388	13	US-10-425-114-22643	Sequence 22643, A
c 36	17	60.7	1668	13	US-10-425-114-22643	Sequence 22643, A
c 37	17	60.7	1725	13	US-10-425-114-22643	Sequence 22643, A
c 38	17	60.7	45450	10	US-09-911-077A-13	Sequence 13, Appli
c 39	17	60.7	175590	10	US-10-371-416-7	Sequence 7, Appli
c 40	16.8	60.0	471	15	US-10-180-158-43	Sequence 43, Appli
c 41	16.8	60.0	811	13	US-10-027-632-152682	Sequence 152682,
c 42	16.8	60.0	811	16	US-10-027-632-152682	Sequence 152682,
c 43	16.8	60.0	1207	13	US-10-045-594D-3	Sequence 3, Appli
c 44	16.8	60.0	1277	8	US-08-722-570-15	Sequence 15, Appli
c 45	16.8	60.0	1277	13	US-10-425-259-16	Sequence 16, Appli

ALIGNMENTS

RESULT 1
US-09-867-701-2002
; Sequence 2002, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2002
; LENGTH: 356
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(356)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-2002

Query Match 64.3%; Score 18; DB 9; Length 356;
Best Local Similarity 80.8%; Pred. No. 87;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 CTGTGGCCAGCTACACAGTACGACT 27

DB 8 CTGTGGCTCAGTAGATAGTACGACT 33

RESULT 2

US-09-918-995-8912/c
; Sequence 8912, Application US/09918995

;; PRIOR APPLICATION NUMBER: JP 2001-204089
;; PRIOR FILING DATE: 2001-05-30
;; PRIOR APPLICATION NUMBER: JP 2001-272697
;; PRIOR FILING DATE: 2001-08-02
;; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 62.9%; Score 17.6; DB 15; Length 9025608;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 CTGTGGCGCAGCTACACAGTACGA 25
DB 2658490 CGGTGGCGCAGCTCCACAGGACGA 2658513

RESULT 7
US-10-101-235A-1
; Sequence 1, Application US/10101235A
; Publication No. US20030182669A1
; GENERAL INFORMATION:
; APPLICANT: Rockman, Howard A.
; APPLICANT: Naga Prasad, Sathyamangla V.
; APPLICANT: Laporte, Stephane A.
; APPLICANT: Barak, Larry S.
; APPLICANT: Caron, Marc G.
; TITLE OF INVENTION: Phosphoinositide 3-Kinase Mediated Inhibition of GPCRs
; FILE REFERENCE: 033072-064
; CURRENT APPLICATION NUMBER: US/10/101.235A
; CURRENT FILING DATE: 2002-03-19
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 627
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-101-235A-1

Query Match 62.1%; Score 17.4; DB 15; Length 627;
Best Local Similarity 77.8%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACACAGTACGACT 27
DB 592 GCTGTGGCGCAGCCTGTCAGGACT 618

RESULT 8
US-10-369-493-43781/c
; Sequence 43781, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374

;; SEQ ID NO 43781
; LENGTH: 1323
; TYPE: DNA
; ORGANISM: NO. US20030233675A1otc punctiforme
US-10-369-493-43781

Query Match 62.1%; Score 17.4; DB 16; Length 1323;
Best Local Similarity 77.8%; Pred. No. 1.7e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACACAGTACGACT 27
DB 379 GCAGTGGCAGCATCTACACATTCGCACT 353

RESULT 9
US-10-282-122A-33322/c
; Sequence 33322, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33322
; LENGTH: 1704
; TYPE: DNA
; ORGANISM: Pseudomonas syringae
US-10-282-122A-33322

Query Match 62.1%; Score 17.4; DB 13; Length 1704;
Best Local Similarity 77.8%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GCTGTGGCGCAGCTACACAGTACGACT 27
DB 797 GCGATGGCGCGCCGACACAGTAAACT 771

RESULT 10
US-10-260-238-2607
; Sequence 2607, Application US/10260238
; Publication No. US20040016025A1
; GENERAL INFORMATION:
; APPLICANT: Budworth, Paul R.
; APPLICANT: Moughamer, Todd G.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION
; FILE REFERENCE: 60111-NP
; CURRENT APPLICATION NUMBER: US/10/260,238
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/325,448
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 6077
; SEQ ID NO 2607
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-260-238-2607
Query Match 62.1%; Score 17.4; DB 16; Length 2000;
Best Local Similarity 77.8%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 1 GCTGTGGCGGAGCTACACAGTACTACT 27
Db 36 GTTGTGCCACAGCTACCAAGTAGGAAT 62

RESULT 11
US-09-954-456-1116/c
; Sequence 1116, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Can
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863

; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1116
; LENGTH: 2726
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-1116
Query Match 62.1%; Score 17.4; DB 9; Length 2726;
Best Local Similarity 77.8%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 2 CTGTGGCGGAGCTACACAGTACTACT 28
Db 1897 CTGAGGCTCAGCTACCAAGTTTGACTC 1871

RESULT 12
US-10-342-887-362/c
; Sequence 362, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 362
; LENGTH: 2726
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-362
Query Match 62.1%; Score 17.4; DB 13; Length 2726;
Best Local Similarity 77.8%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 2 CTGTGGCGGAGCTACACAGTACTACT 28
Db 1897 CTGAGGCTCAGCTACCAAGTTTGACTC 1871

RESULT 13
US-10-172-118-362/c
; Sequence 362, Application US/10172118
; Publication No. US20030224374A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118

```

; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/360,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 362
; LENGTH: 2726
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: D38553
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-362

```

Query Match 62.1%; Score 17.4; DB 13; Length 2726;
Best Local Similarity 77.8%; Pred. No. 1.8e+02;
Matches 21: Conservative 0; Mismatches 6; Indels 0;

Oy

2 CTGTGGCGAGCTACACAGTACGACTC 28
||| ||| ||| ||| ||| ||| ||| |||
D6 1897 CTGAGGTTCAGCTACCAAGTTTGACTC 1871

```

RESULT 14
US-10-171-581-137/c
; Sequence 137, Application US/10171581
; Publication No. US2003010426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; APPLICANT: Mao Mao
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10/171.581
; CURRENT FILING DATE: 2003-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 137
; LENGTH: 2726
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: D38553
; DATABASE ENTRY DATE: 2001-06-18
; US-10-171-581-137

```

Query Match	62.1%	Score 17.4;	DB 15;	Length 2726;
Best Local Similarity	77.8%	Pred. No. 1.8e+02;		
Matches 21;	Conservative	0;	Mismatches 6;	Indels 0;
				Gaps 0;

Qy 2 CTGTGGCGAGTACACAGTACGACTC 28
||| ||| ||| ||| ||| ||| ||| |||
db 1897 CTGAGGCTCAGCTACCAAGTTTGACTC 1871

```

RESULT 15
US-10-084-817-313/c
; Sequence 313, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
; APPLICANT: Jed G. Nuchtern
; APPLICANT: Susan Stuart
; APPLICANT: Sharon E. Plon
; APPLICANT: Jason M. Shohet
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
; FILE REFERENCE: PA-0046 US
; CURRENT APPLICATION NUMBER: US/10/084,817
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 60/270,784
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 365
; SOFTWARE: PERL Program
; SEQ ID NO 313

```

```

; LENGTH: 3079
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID NO. US20030119009A1 333461.2
US-10-084-817-313

```

```
Query Match          62.1%; Score 17.4; DB 15; Length 3079;
Best Local Similarity 77.8%; Pred. No. 1.8e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0;
```

Qy

2 CTGTGGCGCAGCTACACAGTACGACTC 28
|||||
Db 1926 CTGAGGTCAGCTACCAAGTTTGACTC 1900
|||||

Search completed: April 29, 2004, 20:44:18
Job time : 534.667 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 29, 2004, 02:25:04 ; Search time 422.525 seconds
(without alignments)
2872.264 Million cell updates/sec

Title: US-10-624-714-5

Perfect score: 28

Sequence: 1 gctgtggcagctacacagctacgactc 28

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.htg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.om.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htg.mus.*

34: em.htg.pln.*

35: em.htg.rod.*

36: em.htg.mam.*

37: em.htg.vrt.*

38: em.sy.*

39: em.htgo.hum.*

40: em.htgo.mus.*

41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	28	100.0	15178	1	AE006945 Mycobacte
2	28	100.0	324050	1	BX248335 Mycobacte
3	28	100.0	342416	1	BX842573 Mycobacte
4	21.2	75.7	277955	2	AC094488 Rattus no
5	20	71.4	38092	3	AF016429 Caenorhab
6	20	71.4	213045	2	AC120846 Mus muscu
7	20	71.4	220367	2	AC087133 Mus muscu
8	19.8	70.7	110000	2	Continuation (2 of
9	19.6	70.0	890	9	AJ322632 Homo sapi
10	19.6	70.0	157060	2	BX248512 Danio rer
11	19.6	70.0	158106	2	AC128968 Rattus no
12	19.6	70.0	203697	2	AC055819 Mus muscu
13	19.6	70.0	224475	10	AL731774 Mouse DNA
14	19.6	70.0	234545	5	BX470214 Zebrafish
15	19.6	70.0	238330	2	AC106176 Rattus no
16	19.6	70.0	248390	2	AC118791 Rattus no
17	19.6	70.0	259795	2	AC128267 Rattus no
18	19.6	70.0	295543	2	BX539314 Mus muscu
19	19.2	68.6	6513	8	AM19972 Aspergill
20	19	67.9	25461	8	AP004182 Oryza sat
21	19	67.9	110000	2	Continuation (18 o
22	19	67.9	130779	8	AL731638 Oryza sat
23	19	67.9	143961	2	AP005862 Oryza sat
24	19	67.9	144480	8	AP004276 Oryza sat
25	19	67.9	150181	8	AL606604 Oryza sat
26	19	67.9	153005	2	AC011976 Homo sapi
27	19	67.9	153675	2	AP005710 Oryza sat
28	19	67.9	157839	2	AC015718 Homo sapi
29	19	67.9	198305	2	BX640521 Danio rer
30	19	67.9	203375	9	AC083841 Homo sapi
31	19	67.9	203387	2	BX640539 Danio rer
32	19	67.9	227242	10	AC093925 Genomic S
33	19	67.9	224978	2	AC128734 Rattus no
34	19	67.9	266511	2	AC133768 Rattus no
35	18.8	67.1	2042	10	BC016264 Mus muscu
36	18.8	67.1	55010	2	AC107784 Mus muscu
37	18.6	66.4	25893	2	AC014452 Drosophil
38	18.6	66.4	96933	9	AC079621 Homo sapi
39	18.6	66.4	161549	3	AC016157 Drosophil
40	18.6	66.4	165136	3	AC007645 Drosophil
41	18.6	66.4	215899	3	AE003689 Drosophil
42	18.4	65.7	553	9	HS0333123 Homo sapi
43	18.4	65.7	1399	8	AK063465 Oryza sat
44	18.4	65.7	36790	2	AC131448 Strongylo
45	18.4	65.7	61406	2	AC084247 Homo sapi

ALIGNMENTS

RESULT 1
AE006945
LOCUS
DEFINITION Mycobacterium tuberculosis CDC1551, linear BCT 27-APR-2001
complete genome.
ACCESSION AE006945
VERSION AE006945.1 GI:13879900
KEYWORDS
SOURCE Mycobacterium tuberculosis CDC1551
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE
1 (bases 1 to 15178)
AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,

Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J. F., Nelson, W. C., Unayam, L. A., Ermolaeva, M., Khouri, H., Salzberg, S. L., Delcher, A., Uterback, T., Weidman, J., Bhat, H., Gill, J., Mikula, A. and Bhat, W.

Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains

JOURNAL

REFERENCE

AUTHORS

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

TITLE

JOURNAL

FEATURES

source

2 (bases 1 to 15178)
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J. F., Nelson, W. C., Unayam, L. A., Ermolaeva, M., Khouri, H., Salzberg, S. L., Delcher, A., Uterback, T., Weidman, J., Bhat, H., Gill, J., Mikula, A. and Bhat, W.
Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
Unpublished
2 (bases 1 to 15178)
Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J. F., Nelson, W. C., Unayam, L. A., Ermolaeva, M., Khouri, H., Salzberg, S. L., Delcher, A., Uterback, T., Weidman, J., Bhat, H., Gill, J., Mikula, A. and Bhat, W.
Direct Submission
Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
1..15178
/organism="Mycobacterium tuberculosis CDC1551"
/mol_type="genomic DNA"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
198..1457
/gene="MT0401"
198..1457
/gene="MT0401"
/note="similar to GB:L20897 SP:P33221 PID:304887 GB:U00096 PID:1736489; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="phosphoribosylglycinamide formyltransferase 2"
/protein_id="AAK44624.1"
/db_xref="GI:13879901"
/translations="MTDTEOEPTVHERPAAPQDVRVVMGLGSAEPSRELAIAL QGLGASVADGVGAPAHRIADQSVVMTDAEELTAVIRLPPDLVTVTAASVD ALDAVQADECTELVPNARVACTADREGLRLAADQGLTAPDFWVSGELQAV AVHAGFLPVVAGVAGQSSVAGNVEPAWQAGHGVQPOQTGSPVRCASV VEEFLVIMVCSQPNGLIFBFCAPIGHRDAGELESQPOKSTALDAKASIA ARIVKAGVGVSLMNGDEVFADVTVCAGAWTVRSQRLSVFELQARILG LAVDTLMISPGAAVINPHTAGRAVGAAPADALTGALGVPSDDVVFGRGLGVAL ATAPEVAIARERAREVASRLNVDSRE"
1873..3093
/gene="MT0402"
1873..3093
/gene="MT0402"
/note="similar to SP:P55218; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="O-succinylhomoserine sulphydrolase"
/protein_id="AAK44624.1"
/db_xref="GI:13879902"
/translation="MTDSESVRTPKALPDGVQATVVRGGMRLSGFEETAEAMYLTS GYVGSAAVAEKSFAGELDHVYVRYGNTVSVFEERLRIEGAAPATASGMAAVF TSLGALLGAGDLRAVARSFGSCFVVCSEILPRWGVTQVFDGDLDSOMERLVSPTO AVFPEPSPMOSLVDIAAVTLAHAAGAKVLDNVFATPLIQQGFPLGVDDVYVSGT KHDGGRVLGAILGDRIYIDGPKMRHTGPAWAFNAWLLKGLLETLAIRVOHS NASAQIAEFLNGHPSVRWRYPLFSPHQDLAKRQMSGGTGVTFALDCPDVAKQ RAFEVLDRKRLIDISNNLGDASKSLVTHPATTHTRAMPGEGRRAAIGLGDGVVRSVGLS DTDLLADIADRLAS"
complement (3090..4502)
/gene="MT0403"
complement (3090..4502)
/gene="MT0403"
/note="similar to GP:2708705; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="NADH dehydrogenase"
/protein_id="AAK44625.1"
/db_xref="GI:13879903"
/translation="MTLSGGEPSAVGGRHVRVVIIGSGFGLNAKALKRADVDITLIS

KTTHLFQPLLYQVATGILSEGGDIAPTTRILRLRQKVRVLLGEVNAIDLKAQVTSK LMDHTRRTTFFVVGAGTGVGAQIVELAERTLAGAFRTITPSECRVILLDAAP AVLPMPKGLKQAKLEKMDVEVOLNAMVTAVDKGITIKEDGGERIECAKVM AAGVAAPLGMIAEGSDTEIDRAGRIVVEPDLTVKHPNVFVGOLMFVPGVGA QCAIQGARYATVTKHMGNDPANRPPHFNKSGMATISRSASVAQVCKLEFAGY FAWLAWLVHLVLYGNRIIAALFANGCISPMGARGQMAITSMIARLVMTLMEQQ AQGLAARAEQAEHAEQERAG"
4644..5969
/gene="MT0404"
4644..5969
/gene="MT0404"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44626.1"
/db_xref="GI:13879904"
/translation="WAVGRCAIPRFDOAASGAINGGVHLSDGSTSPARQLPAPWPG DAGAAAGRAGVCCRCNGLPHVSDVGYSHRFDHPAGVAGGCPAGAGAGLAVDDPG QLAALIRIVADPDPAVOURERARDREVSINWADGMEVYQALYATDAQALDARL NALVATVCGDPRSTDQRDALGALAGADRACRCNDPCARAGRPVSAVVHVVV EQASVGHGQAPAAALGDLIPAEVLAEAKTAGLQPIVPVAGTEPGYSPVSKLAAP VRADLTCTRAPGCDRPTAQCDLDTIAFDGAGTHAANLKLCLHLHLATFCGWRQO QLPDGTIVITLPGNQITVTFGSALLFPALCTPTGDPPEPARADRRGORTAMPPRR ASTRTQNRACIAAERHNRHQARRIAQAQAAVIATETHGPPDPDDPPPPF"
complement (5985..6704)
/gene="MT0404.1"
complement (5985..6704)
/gene="MT0404.1"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44627.1"
/db_xref="GI:13879905"
/translation="MTEPRFVAVVISAGLSAIPMVGGPLOTVFDIAEETRHAEFT TRECEVGGADTVLSIDKNPELPLLSQIAEAATTSMEAKRRLIAQAAALBDD QKVPASLIATLSQEPVHIHALVRLAKAKSPDODETORREVWRAASKVFPVPL AALITQGVATATVWNGTGTTPAESGHILIHDSDFGHRLLALYRAADAGAEILLI LPSGGSAPTGDHPTPHPTSR"
6803..7207
/gene="MT0405"
6803..7207
/gene="MT0405"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44628.1"
/db_xref="GI:13879906"
/translation="MDMPLGDIYETFRHWSGKPRAMQPSGWRWFGKIVDGLCEV LDEHLAVRRGVPAACVCEWLSSEAVAETLLALSACVVIDKGTSPSRRLRNPDKGF PNVALLRDMAPSEHGRSCSSARGRLCLMS"
7189..7605
/gene="MT0406"
7189..7605
/gene="MT0406"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44629.1"
/db_xref="GI:13879907"
/translation="MPEHGLPVRLGWLREDRKPLLNKLLVLGLHNLNVDPDNGY GEVLDPEFRTVWMSANTVRAGSLHYGVFACDDPTLVEEATAFVADVIXSSEIDT TCAGPEXNLVQVFEFDAAAEAMEENAEPPDDGDEW"
complement (7646..8116)
/gene="MT0407"
complement (7646..8116)
/gene="MT0407"
/note="identified by Glimmer2; putative"
/codon_start=1

```
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44630.1"
/db_xref="GI:113879908"
/tanslations="MKYNNRTCLXPVPVWAGAYRLGRRVVVGIGWVWSLGGCC
GFLGDHGVNPAJLVVAVFGDDVVGAVLRSGAGPARHHRGALPAAVGAGVGSRWYTG
GGRQREQQAAGGIGLGLVTRPNHAGVQLLAPAPEGGGQMAPIRRTF"
8257.8505
/gene="MT0407.1"
/genes="MT0407.1"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44631.1"
/db_xref="GI:113879909"
/tanslations="MHALLVGLAILTAIPIAVLIGSSPAHADTDIGOPCSPEGAKL
WGNPGPIYCBRTADGQLQWVSIPAWALCVAFCDRPGP"
complement(8542..9183)
/gene="MT0408"
/genes="MT0408"
complement(8542..9183)
/gene="MT0409"
/genes="MT0409"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK44632.1"
/db_xref="GI:113879910"
/tanslations="MGVIARVWVGAAGSLAVLAAAPTAGAEFTGALPMTSSGSGP
VIGDAAALRQISQQLFSGDPTVOEDGSDAAQFTTAAAVADRDVASVFLPQRY
LGQCNFAGSGAGFGARVRRDTGQMGAMLVVAKSTVSDVDAKACVXGWRKATAG
TPTSMCNWNTYPPFADTRGEGYFVLLAGTASDFCSAPNANVRTTASWPG"
complement(9190..10419)
/gene="MT0409"
/genes="MT0409"
complement(9190..10419)
/gene="MT0409"
/notes="identified by match to Prosite PS00013"
/codon_start=1
/transl_table=11
/product="lipoprotein, putative"
Query Match 100.0%; Score 28; DB 1; Length 15178;
Best Local Similarity 100.0%; Pred. No. 0.0072;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCTGTGGCGGAGCTACACAGTACGACTC 28
Db 4523 GCTGTGGCGGAGCTACACAGTACGACTC 4550
|||||
|||||

RESULT 2
BX248335 324050 bp DNA linear BCT 11-JUN-2003
LOCUS
DEFINITION Mycobacterium bovis subsp. bovis AF2122/97 complete genome; segment
2/14.
ACCESSION BX248335 BX248333
VERSION BX248335.1 GI:31617046
KEYWORDS complete genome.
SOURCE Mycobacterium bovis subsp. bovis AF2122/97
ORGANISM Mycobacterium bovis subsp. bovis AF2122/97
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
1
REFERENCE
AUTHORS
Garnier,T., Eiglmeier,K., Camus,J.-C., Medina,N., Mansoor,H.,
Pryor,M., Duthoy,S., Grondin,S., Lacroix,C., Monsemp,C., Simon,S.,
Harris,B., Atkin,R., Doggett,J., Mayes,R., Keating,L.,
Wheeler,P.R., Parkhill,J., Barrall,B.G., Cole,S.T., Gordon,S.V. and
Hewinson,G.
The complete genome sequence of Mycobacterium bovis
Online Publication
PNAS 10.1073/pnas.1130426100 ( Microbiology )
```

```
REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 324050)
Garnier,T.
Direct Submission
Submitted (24-MAR-2003) Garnier T., Unite de Genetique Moleculaire
Bacterienne Institut Pasteur 28,rue du Dr Roux 75724 PARIS cedex
15, France. e-mail:tgarnier@pasteur.fr Submitted on behalf of the
Mycobacterium bovis sequencing teams, TB Research Group, Veterinary
Laboratories Agency Weybridge, Woodham Lane, New Haw,Addlestone,
Surrey KT15 3NB, UK. Sanger Centre, Wellcome Trust Genome Campus,
Hinxton,Cambridge CB10 1SA, UK. P4 Annotation, Genopole, Institut
Pasteur, 28 Rue du Docteur Roux, 75724 Paris Cedex 15, France.
Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28
rue du Docteur Roux, 75724 Paris Cedex 15, France
location/Qualifiers
1..324050
/organism="Mycobacterium bovis subsp. bovis AF2122/97"
/mol_type="genomic DNA"
/strain="AF2122/97"
/db_xref="taxon:233413"
132..2027
/locus_tag="Mb0290"
132..2027
/locus_tag="Mb0290"
/notes="Mb0290, -, len: 631 aa. Equivalent to RV0282, len:
631 aa, from Mycobacterium tuberculosis strain H37RV,
(99.8% identity in 631 aa overlap). Conserved hypothetical
protein, similar to Y14967|MCB628.18c hypothetical
protein from Mycobacterium leprae (573 aa), FASTA scores:
opt: 916, E(): 0, (38.7% identity in 568 aa overlap). Also
similar to Mycobacterium tuberculosis proteins e.g.
Z94121|MTY15F10.26 (619 aa), FASTA scores: opt: 743, E():
0, (29.9% identity in 612 aa overlap). Member of CBXP,
CBXP family - 9 members in Mycobacterium tuberculosis.
Contains PS00017 ATP/GTP-binding site motif A (P-loop)."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD93154.1"
/db_xref="GI:31617047"
/translation="MAGVGAGDSGGVERDDIGWVAASPVASRVNGKVDADVVGREATC
CRALGIYQKRPDPDLAAASGFAALTRVAHDQCDANTGLAAGDQSIGVLEAASRT
ATTAGLQRLQVEALNALGFLYDTGLYLRFRATGPDHFLVAALASTGGPEEFAKA
NHVSGITERRAGWEARLAVINRYRAERSDVVKLLTPMVDPLDEAFSHAAKIT
LGTALARLGMFAPALSYLEEPGPVAAVADGALAKALVLAHVDEESASEVLODLYA
HEPENEQVQALSDFSGIVTTTACRIEARTDPMDPATEPGAEDVDPAAHERKALL
HEARLQAEFGLDEVRQVSELKSSVAMELVKRGKLTVAQTHLVFAGPPGTKT
TIARVAKIYGLGLLKRNIKEVHRADLIQGHIGETAKNALIDSALDGLFLDEA
YALVATKANDFGLVAIDTLTARMENRDLRVIIAGTRADLDKFLDNEGRLSRPTR
NIDFPTYSHELVEIAHVAQRDSVFQSSALHDLFAKLAESTDPTNGISRRSL
DIAGNRFVRNIVERSEBERFRLDSEHAGSGEFSDEBLMTITADVGSRVPEPLRG
LGLSVRA"
2024..3640
/locus_tag="Mb0291"
2024..3640
/locus_tag="Mb0291"
/notes="Mb0291, -, len: 538 aa. Equivalent to RV0283, len:
538 aa, from Mycobacterium tuberculosis strain H37RV,
(100.0% identity in 538 aa overlap). Possible conserved
membrane protein, similar to several hypothetical
mycobacterial proteins e.g. Z94121|MTY15P10.16|rv3895c
from Mycobacterium tuberculosis (495 aa), FASTA scores:
opt: 698, E(): 0, (37.6% identity in 492 aa overlap);
rv1782; rv3450c; rv3869; and Y14967|MCB628.16|MCB628.17c
from Mycobacterium leprae (481 aa), FASTA scores: opt:
672, E(): 1.5e-31, (37.2% identity in 506 aa overlap).
Contains PS00017 ATP/GTP-binding site motif A (P-loop)."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="POSSIBLE CONSERVED MEMBRANE PROTEIN"
/protein_id="CAD93155.1"
/db_xref="GI:31617048"
```


2 Camus, J.C., Pryor, M.J., Medigue, C. and Cole, S.T.
Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv
Microbiology (Reading, Engl.) 148 (Pt 10), 2967-2973 (2002)
2255591
12368430
3 (bases 1 to 342416)
Parkhill, J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the *Mycobacterium tuberculosis* sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 7524 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On or before Nov 21, 2003 this sequence version replaced
gi:3261524, gi:3261526, gi:3261527, gi:3261529, gi:3242298,
gi:3261606, gi:3261689, gi:3261708, gi:3261703, gi:3261760,
gi:3261781, gi:3261797, gi:3261800, gi:3261825, gi:3261837.
Notes:
Details of *M. tuberculosis* sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/).
Location/Qualifiers
1. .342416
/organism="Mycobacterium tuberculosis H37Rv"
/mol_type="genomic DNA"
/strain="H37Rv"
/db_xref="taxon:83332"
223. .2118
/locus_tag="Rv0282"
223. .2118
/locus_tag="Rv0282"
/function="UNKNOWN"
/note="Rv0282, (MTV035.10), len: 631 aa. Conserved hypothetical protein, similar to Y14967/MLCB628.18c hypothetical protein from *Mycobacterium leprae* (573 aa), FASTA scores: opt: 916, E(): 0, (38.7% identity in 568 aa overlap). Also similar to *Mycobacterium tuberculosis* proteins e.g. Z94121/MTY15F10.26 (619 aa), FASTA scores: opt: 743, E(): 0, (29.9% identity in 612 aa overlap). Member of CFXQ, CBXP family - 9 members in *Mycobacterium tuberculosis*. Contains PS00017 ATP/GTP-binding site motif A (P-loop)."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAA17357.1"
/db_xref="GI:2909471"
/translation="MAGVSGSGGVVERDDIGMVAASPVASRVNGVKVDADVGRFATC CRAIGIAYQRKEPPDLAARGSFALTRVAHQDCDAWTGLAAGQSIGVLEASRT ATTAGVLCORVELADNALGFLYDTGLYFRFRATGDDFFHLAYAAALASTGPEEFKA NHVVSIGITERAGWRARMLAVINVRAEWSDVVKLLTPMNDPDLDAFASHAAKIT LGTALARLGNFALSIVLEPDGPVAAVADGAKALVLRHVDSAESVLDLYA AHENEQVQALSDFIGVITTAGTAREAKTDPDWATERCAEDFVDPRAHERKAALL HEALQAEIGLDEKVRQSVSMELVRKQGLTVQAQTHLHVFAGPGTGKT TIARVAKYCGGLGKRENIREVRADLIGCHIGTEAKNTAIIDSLDGLVLFDEA YALVATKANDFGVLVADITLLARMENDRLVVIAGYRADLDFLDTWEGLSRFRTR NIDPPSYSHELVIAHKMAEQDSVFQSQALHDLALFAKLAESTPTDPTNGISRSL DAGNGRFVNIVERSEEREFRDLHSHAGSGCFSDLELMTITADVGSRVPELLRG LGLSVRA"
1375. .1398
/locus_tag="Rv0282"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
2115. .3731
/locus_tag="Rv0283"
2115. .3731
/locus_tag="Rv0283"
/function="UNKNOWN"
/note="Rv0283, (MTV035.11), len: 538 aa. Possible conserved membrane protein, similar to several hypothetical mycobacterial proteins e.g.

294121/MTY15F10.16|Rv3895c from *Mycobacterium tuberculosis* (495 aa), FASTA scores: opt: 698, E(): 0, (37.6% identity in 492 aa overlap); Rv1782; Rv3450c; Rv3869; and Y14967/MLCB628.16|MLCB628.17c from *Mycobacterium leprae* (481 aa), FASTA scores: opt: 672, E(): 1.5e-31, (37.2% identity in 506 aa overlap). Contains PS00017 ATP/GTP-binding site motif A (P-loop)."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="POSSIBLE CONSERVED MEMBRANE PROTEIN"
/protein_id="CAA17358.1"
/db_xref="GI:2909472"
/translation="MTNQDHDDHDDHRRSPASTPVNNNPDKVYRRGFVHQVTVG WRFYMRRIAAGIALHVRLEQMLVPLRTQSRALVGLVITGLIGSFVSLIRNGQAG SNAVLDRSSTAALVVRVEOLHPVLNLTARLIVGRFVPTVKTSDELDPGRNLGIG IPGAPERKQNTSDANWTVCDELNAPSRGDAGGVTVIAGPLEDGAARALGPGQ AVLVDGAGTWLLMDGKRSPIDLDADHAVITSGLGCAGVAPAPRIIASGLFNAIPAPPL TAPIIDAGNPASFGVPAIGAVVSYALKDSGKTIISDTVOIYAVLPDGLQGISFVLA AILNNNSYGLQQPQPRILGADAVAKLPSVRLDTRRYPSEFVSLVDTRDFVTCAYWSK PVGAATSSLTLAGSALPVPDAVHTVELVGAGGVATRVALAAGTGYFTQTGGGPD APGAGSLFWSDTGVRYGIDNEPGVAGGKAVEALGLNPPVPPIPWVLSLFPVGPPT LSRDALLAHDLTLPDSRPAPVSAEGYR"
3024. .3047
/locus_tag="Rv0283"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
3728. .7720
/locus_tag="Rv0284"
3728. .7720
/locus_tag="Rv0284"
/function="UNKNOWN"
/note="Rv0284, (MTV035.12), len: 1330 aa. Possible conserved membrane protein, similar to products of two adjacent *Mycobacterium leprae* genes, MLCB628.16c (744 aa) and MLCB628.15c (597 aa); and throughout its length to several large *Mycobacterium tuberculosis* proteins: Rv3447c, Rv3870, Rv1784, etc. Y14967/MLCB628.15 (744 aa), FASTA scores: opt: 942, E(): 0, (33.8% identity in 730 aa overlap); Y14967/MLCB628.14 (597 aa), FASTA scores: opt: 613, E(): 3.1e-30, (31.7% identity in 615 aa overlap); 294121/MTY15F10.17 (1396 aa), FASTA scores: opt: 2.2e-32, (35.4% identity in 1321 aa overlap); 295389/MTY17.19 (1236 aa), FASTA scores: opt: 652, E(): 2.2e-32, (35.4% identity in 1321 aa overlap). Contains three PS00017 ATP/GTP-binding site motif A (P-loop)."
/codon_start=1
/evidence=experimental
/transl_table=11
/product="POSSIBLE CONSERVED MEMBRANE PROTEIN"
/protein_id="CAA17359.1"
/db_xref="GI:2909473"
/translation="MSRLIFARRRLAPSPSHOQTIIIEAPPELPRVIPPSSLRRALP YLIGLIVGMIVAVATQMRVISPOTLFFPFLVLLAATLYRGNDKKRTEVDAERA DLYRLVSVNRNIRAQAASQASALWSHPDPTALASVPSRRQWRDHPDPLVLA RGHVPLVTRNDITLRVNDTADIEIPVSHSALSLDQTSIGDPTGIDLTQVSPITVL GERAVRAVLRAWIAQAVTWDPTVLGVALAARDLEGRDNWMLKPLHVDIIPGRDLAL GPARNLSDPDELLIALLGVLDARPAFTQPTDARLHLLIVVDDPYDLGASPLAVGR AGTVVHDSASAPHEQSDDEKPLIRVAHGAIERWQGGQPYDDAQSADAAH LARRVHDSNPTHAGLSAATRGASFTLLIGIESADRSDQTLMSILLSTLTHSAERLIVI GVTGGEPLMDLDEAGCGPHGIMTGMGSKSQTLMSILLSTLTHSAERLIVI YADPKEAGADSFRDPQVAVINNAEKSLADPADTLPGEVAREMLLEAGRKV OGSAPNVSLEYENAIAGHSPLPIPTLFVADDEFTMLADHPEYAEFLDYVARKGSF RTHILPQSDTLDVGKIDIDKNTAVRIGLVKVASPSVSKQIIGVEDAYHIESGHEKGV GLIVAPAGATPIRFRSTYVDGIYEPPQAKAVVQSVPEPKLFTAAVEPDGTVIAD TDEQPADPPRKLTIATIGQLARYGPRAPOLMLPDELDTIPLSALAARAGVGRQWRW PLGEIDRPFEMERDPLVDFARSAGNMVHGGFKSGKSTALQTFILSASLSHREVS FYCLDYGQQLALQDLAHVGSVASALEPFRIRTFGELEQLLRQREVRFDKGAN GTCDDGFGCEFLVLDNYLGFGRDNTQDNTRNPLLARVLELVNLYGLAYHVIITTP SWLEVPALARVDGLRLRLHDAADSNRVVVGALRPADAVPHDQPGRLTMAAEHF LFAPELDAQTPVAIINARYPGMAAPVRLPTNLPAVGLRGVGDOLVGOREEE DLAPVILDAAANPLMWGDAESGKTLRHIRTVRHSTADRVAFVTRDRRLHDDV EFLDFDNETANIDRIIPAMGLANLIEARRPPAGMSAELSRFTAGHTVHLIIDDV DQVPDPSAMTGPYIGRIPLGLLAQAGDGLGRVITVGRATGSAHLLMTSFLRRF

* consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 277955: contig of 277955 bp in length.

FEATURES
 source
 1. 277955
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-4G21"
 1. 1208
 /note="wgs end extension
 clone_end:Sp6"
 4369. 5099
 /note="clone boundary
 clone_end:Sp6
 site:EcoRI
 end sequence:BH306443"
 270424. 271098
 /note="clone boundary
 clone_end:T7
 site:EcoRI
 end sequence:BH306441"
 276585. 277955
 /note="wgs end extension
 clone_end:T7"

misc_feature
 276585. 277955
 /note="wgs end extension
 clone_end:T7"

misc_feature
 270424. 271098
 /note="clone boundary
 clone_end:T7
 site:EcoRI
 end sequence:BH306441"

misc_feature
 276585. 277955
 /note="wgs end extension
 clone_end:T7"

ORIGIN
 Query Match 75.7%; Score 21.2; DB 2; Length 277955;
 Best Local Similarity 88.5%; Pred. No. 19;
 Matches 23; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 TGTGGCGCAGCTACAGTACTC 28
 |||||
 Db 188713 TGTGGCATCTACAGTACTC 188738

RESULT 5
 AF016429 38092 bp DNA linear INV 04-OCT-2003
 LOCUS AF016429 Caenorhabditis elegans cosmid T21H3, complete sequence.
 ACCESSION AF016429
 VERSION AF016429.1 GI:2291245
 KEYWORDS HTG.
 SOURCE Caenorhabditis elegans
 ORGANISM Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 38092)
 Wilson, R.
 Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 98069613
 9851916
 2 (bases 1 to 38092)
 Scheet, P.
 The sequence of C. elegans cosmid T21H3
 Unpublished (2001)
 3 (bases 1 to 38092)
 Waterston, R.
 Direct Submission
 Submitted (29-JUL-1997) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 4 (bases 1 to 38092)
 Waterston, R.
 Direct Submission

Submitted (05-OCT-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 5 (bases 1 to 38092)
 Waterston, R.
 Direct Submission
 Submitted (23-MAY-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 6 (bases 1 to 38092)
 Waterston, R.
 Direct Submission
 Submitted (21-NOV-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 38092)
 Wilson, R.
 Direct Submission
 Submitted (04-OCT-2003) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 email: submissions@watson.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one ml3 subclone.

For a graphical representation of this clone sequence and its
 analysis see:
 http://www.wormbase.org/db/seq/sequence?name=T21H3;class=Sequence

NEIGHBORING CLONE INFORMATION

The 5' clone is W0788, 200 bp overlap; the 3' clone is F25A2, 300
 bp overlap. Actual start of this clone is at base position 197 of
 T21H3; actual end is at 37992 of T21H3.

NOTES:
 Coding sequences below are the result of integration and manual
 review of the following data: computer analysis using the program
 Genefinder (P. Green and L. Hillier, personal communication), the
 large scale EST projects of Yuji Kohara
 (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C.
 elegans ORFome cloning project (http://wofdb.dfci.harvard.edu/),
 similarity to other proteins from BlastX analyses
 (http://blast.wustl.edu/), sequence conservation with C. briggsae
 using Jim Kent's WABA alignment program (Genome Research
 10:1115-1125, 2000), individual C. elegans Genbank submissions,
 and personal communications with C. elegans researchers. TRNAS
 are predicted using the program tRNAscan-SE (Lowe, T.M. and
 Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

Location/Qualifiers
 1. 38092
 /organism="Caenorhabditis elegans"
 /mol_type="genomic DNA"
 /strain="Bristol N2"
 /db_xref="taxon:6239"
 /chromosome="V"

FEATURES
 source

```
gene
CDS
/codon_start=1
/product="Hypothetical protein T21H3.4"
/protein_id="AAB65365.2"
/db_xref="GI:37515214"
/db_xref="WormBase:T21H3.4"
/translation="MYLILKILYILWLTFTTSAGSIVHTQNSLNGAKFYEFRRSPQ
KIRDFRIRKASGNFTPTVICDSQIGIALLNENPADIRMFADSHHDPDIRRPSLELD
POFVTRIDGCGVRRAYGVPSGFTVYSIRLIVSHNSDELTEFDKIYDVTCSLHLKTM
AVETSYDIISPOTTTISSTVHSRAGPKCKYSLHHTVRGGPRTVSAVGEVKKAT
KXINGKSLALHVMKNEIPNSGNFPVKYKPHRDSLCYRPSRLNYSKYSGRSNKTIN
TDSMHNFOYTHRWKCAELYENNRQMPMLIFKYACVVDHNNRTYAIIDDNGCSL
DEELIPEYDIENGVIYTPSKAFRANSNHHFKCMISVCSAVDPSCRNSVPPKCKV
GLAKKRGQPEEMTIEQRLIRHELMKVRNITVEVGRPTKLSHGAVLPEVEEN
SLNVNREYIEKLEHILRSYQIWTWILCVLNVFLATVLCVMAFRVLKKNKYIMDIKDK
TKTVRRVVP"
/sequence="16427"
/gene="cmd-1"
/join(15543..15576,15858..16096,16194..16341,16399..16427)
/gene="cmd-1"
/standard_name="T21H3.3"
/notes="contains similarity to Pfam domain PF00036 (EF
hand); coded for by the following C. elegans cDNAs:
CESRA05F, AJ132193, YK90H7.3, YK90H7.5, YK201A12.3,
YK201A12.5, YK201F1.5, YK206E8.5, YK283C1.3, YK283C1.5,
YK326A8.5, YK326E5.5, YK338H9.5, YK339B9.5, YK371B12.5,
YK457E3.3, YK457E3.5, YK471H6.5, YK478B5.5, YK494F9.3,
YK494F9.5, YK497B5.3, YK497B5.5, YK499D8.5, YK501F5.3,
YK501F5.5, YK513F1.5, YK516E10.5, YK534D11.3, YK534D11.5,
YK560G12.3, YK560G12.5, YK574D3.5, YK586G4.5, YK601B8.5,
YK602A7.3, YK602A7.5, YK623H4.5, YK638C4.3, YK638C4.5,
YK403B3.3, YK640B3.5, YK653E1.5, YK598I0.5, YK660H11.3,
YK660H11.5, YK661B7.5, YK661B7.5, YK747I2.3, YK817H07.3,
YK817H07.5, YK851G11.5, YK900C01.5, YK1068F03.5,
YK1102F09.5, YK1227E07.3, YK1306D04.5"
/codon_start=1
/product="Calmodulin protein 1"
/protein_id="AAB65364.1"
/db_xref="GI:2291247"
/db_xref="WormBase:T21H3.3"
/translation="MADQTEQIAEFAKSLFDKDGDTITTKELGTWMSLQGNP
TEALQDMINEVDAGNDITDFPFTLTMARKGDTDSSEIREAFRVDKDGNGFIS
AABLHVMTNLGKLTDEEDIREADIDGQVNYEYFVTMTTK"
/complement(17198..20788)
/gene="T21H3.5"
/complement(join(17198..17297,17490..17621,18777..19108,
19677..19878,20403..20788))
/gene="T21H3.5"
/standard_name="T21H3.5"
/notes="contains similarity to Leptospira interrogans
Conserved hypothetical protein: TR:Q8P0F1; coded for by
the following C. elegans cDNAs: YK497A7.5, YK513F1.3,
YK623H4.3, YK804C11.3, YK1068F03.3, YK1306D04.3"
/codon_start=1
/product="Hypothetical protein T21H3.5"
/protein_id="AAB65366.1"
/db_xref="GI:2291249"
/db_xref="WormBase:T21H3.5"
/translation="MYQTEPTLTAAVQIILTWGCAISLIYVNVLLIYSINERTDN
RFRWFPTILMYLCPNALAVLPVHPFIFFARGLIINVFPPARILNQLVQKIVY
GLVIGITQMCIAPIVLVRLMVKKRSICMTLHYTKIRKULFASLFLPITLBIYL
IFFTPNOKISYSEVSHDGVKSLGSHDFSYFVDMWSLLYFTTFFITFYIVSLV
FKFYKSTRNYSNEIDGVSHSINRILSIALONLISIPHFVITLISLKW
LGVRLPASPGGCTITLTITSLTIFVNFRAVRLQGLFIIDFFRQKNVLSVFPVNDNQ
```

```
gene
CDS
QADQLQMTVGAPISMDYRDGGGGGLVKGNGIKLLKDFGDFDFYNNIK"
29364..34907
/gene="ptr-16"
/join(29364..29522,29571..29767,30237..31671,32317..32814,
33823..33946,34000..34044,34231..34280,34770..34907)
/gene="ptr-16"
/standard_name="T21H3.2"
/notes="contains similarity to Pfam domain PF02460 (Patched
family)"
/codon_start=1
/product="Patched related family protein 16"
/protein_id="AAB65363.1"
/db_xref="GI:2291246"
/db_xref="WormBase:T21H3.2"
/translation="MDNLLRRDYLDGVDLSLAELYQYNFTVPCSGVRGENKLECSFSDL
CSGQNDNQYLNKLRNNEKNQYDIFLNLIRYNSARLHPRLTYPTMHLVNDY
YVGEHFAVKIDPNTNIVSRVKKVVLVPRDQRTQVSVQVNNWETSLFDVNFDPH
FLNMTVNSDAMAREVTRNGMTQVPPFSFSAVAVIFIRATNSREHFVSHNVMAIL
GIAGPLMATGTAFGFLFLGPPNSIILWPFLLIGVCGDDVFIILHMKTKDSSEL
EDRIATMEERAGPSITVTSATNCLSFALGATPAPLSFLCLYTCVAGVDFVYQLTF
FVALVYEEKLMKLVGEEKIEAAMRRPKQVLSIQNSIRSTAGAHPPANPNGI
VSYRFLDKWTKLCLLLVLCGYTASYYGCKTMEIKMDTINLIMDSPLNHIAWTY
ERFLMSGQLVMVFNPNPDLTGKSENLEHVELVNRPETLPYSMGKNSITWLSFYL
QSSLYHNKGFNGLLQWLEDSGGGARWMDMLKKDENGTLIGIDKFFAFACAMG
DDANWSTREKLQKQGVGAHEHFNVTVFSYSFYIDQLDSIGATMTSIVAAITM"
Query Match 71.4%; Score 20; DB 3; Length 38092;
Best Local Similarity 82.1%; Pred. No. 75;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 GCTGGGGCGAGCTACACAGTACGACTC 28
DB 36931 GCTGCAGTGCAGCTACACAGTACGCTCGC 36958
RESULT 6
AC120846
LOCUS
AC120846
DEFINITION
Mus musculus chromosome 7 clone RP24-317L6 map 7, linear
PROGRESS ***, 12 unordered pieces.
AC120846
AC120846.6 GI:40217736
VERSION
HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus;
Birren,B., Nussbaum,C. and Lander,E.
1 (bases 1 to 213045)
Mus musculus chromosome 7, clone RP24-317L6
Unpublished
REFERENCE
2 (bases 1 to 213045)
Biren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouhgalier,B., Brown,A., Camarato,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczyk,J., Levine,K., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
```